

MILENA TÁPIA

**REDES NEURAIS ARTIFICIAIS: UMA
APLICAÇÃO NA PREVISÃO DE PREÇOS DE
OVOS**

FLORIANÓPOLIS - SC

2000

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA
COMPUTAÇÃO**

Milena Tápia

**REDES NEURAIIS ARTIFICIAIS: UMA
APLICAÇÃO NA PREVISÃO DE PREÇOS DE
OVOS.**

Dissertação submetida à Universidade Federal de Santa Catarina como parte dos
requisitos para a obtenção do grau de Mestre em Ciências da Computação.

Orientador:

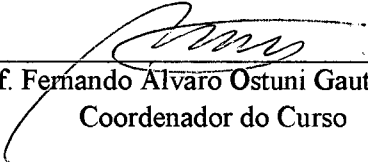
Prof. Mauro Roisenberg, Dr.

FLORIANÓPOLIS, DEZEMBRO DE 2000.

REDES NEURAIS ARTIFICIAIS: UMA APLICAÇÃO NA PREVISÃO DE PREÇOS DE OVOS

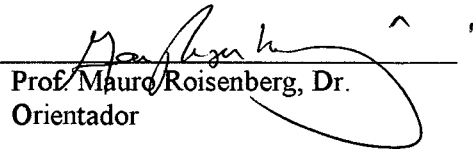
Milena Tápia

Esta Dissertação foi julgada adequada para a obtenção do título de Mestre em Ciência da Computação Área de Concentração Sistemas de Conhecimento e aprovada em sua forma final pelo Programa de Pós-Graduação em Ciência da Computação.

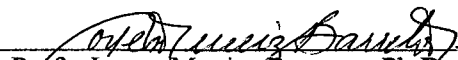


Prof. Fernando Alvaro Ostuni Gauthier, Dr.
Coordenador do Curso

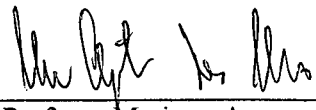
BANCA EXAMINADORA




Prof. Mauro Roisenberg, Dr.
Orientador



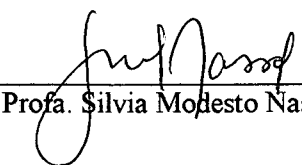
Prof. Jorge Muniz Barreto, Ph.D.
Professor Coorientador



Profa. Maria Augusta Soares
Machado, Dra.



Prof. Pedro Alberto Barbata, Dr.



Profa. Silvia Modesto Nassar, Dra.

PUBLICAÇÕES

TÁPIA, M.; ROISENBERG, M.; BARETTO, J.M. Aplicação de Redes Neurais Diretas na Previsão de Preços Agrícolas: Um Exercício Empírico. In I Workshop da Pós-Graduação em Ciências da Computação, Florianópolis, SC, Maio, 2000.

TÁPIA, M.; ROISENBERG, M.; BARETTO, J.M. Aplicação de Redes Neurais Diretas na Previsão de Preços de Ovos. In InfoAgro 2000 - Congresso e Mostra de Agroinformática, Ponta Grossa, PR, Outubro 2000.

A Deus
*"... porque a ele nenhuma
coisa é impossível ..."*
Lucas 1,37.

A meus pais, Pedro e Iraci Tápia
pelo apoio constante e em
todos os sentidos.

Agradecimentos

À Universidade Federal de Santa Catarina;

À Universidade do Contestado;

Aos professores do Curso de Pós-Graduação
em Ciências da Computação, em especial
ao Prof. Pedro Barbeta;

Ao orientador Prof. Mauro Roisenberg,
pelo acompanhamento pontual e competente, pela
amizade, incentivo e principalmente pela paciência...

Ao Prof. Coorientador, Jorge M. Barreto,
pelos conhecimentos transmitidos;

A minha amiga irmã Alessandra,
pela amizade, coragem, fé e
experiência compartilhada;

Aos meus amigos, Roberto e Denise
pelo apoio, carinho e amizade
em todos os momentos;

Ao meu amigo José Gonçalo,
pela sua capacidade de acolher, ouvir,
aceitar e entender aquilo que a outros ouvidos e
corações parecia inaceitável e impossível;

A família GOU,
pela acolhida, amizade, companheirismo,
força, oração e incentivo constantes;

Aos colegas e amigos que colaboraram,
às vezes até sem saber, gerando um ambiente
positivo e favorável para a realização desta pesquisa;

Aos meus pais,
que acreditaram em meus sonhos
e providenciaram tudo para a realização deles;

E acima de tudo, a Deus,
meu autor, princípio e fim.

*"If we could first know where we are,
and whither we are tending,
we could better judge what to do,
and how to do it."*

Abraham Lincoln

SUMÁRIO

SUMÁRIO	VI
LISTA DE FIGURAS	IX
LISTA DE TABELAS	XI
LISTA DE QUADROS	XII
LISTA DE REDUÇÕES	XIII
RESUMO	XIV
ABSTRACT	XV
CAPÍTULO 1 - INTRODUÇÃO	01
1.1 NATUREZA DO PROBLEMA	02
1.2 OBJETIVOS DA PESQUISA	05
1.2.1 Objetivo geral	05
1.2.2 Objetivos específicos	05
1.3 ORGANIZAÇÃO DO TRABALHO	06
CAPÍTULO 2 - MÉTODOS ESTATÍSTICOS DE PREVISÃO	07
2.1 ANÁLISE DE SÉRIES TEMPORAIS	08
2.1.1 Conceitos básicos	08
2.1.2 Objetivos	09
2.1.3 Previsão de séries temporais	10
2.1.4 Medidas de desempenho	12
2.2 CONSIDERAÇÕES FINAIS	15
CAPÍTULO 3 - METODOLOGIA DE BOX & JENKINS PARA MODELOS	
UNIVARIADOS	16
3.1 INTRODUÇÃO	16
3.2 MODELOS POSSÍVEIS DENTRO DA METODOLOGIA BOX & JENKINS.....	17
3.2.1 Modelos estacionários	17
3.2.1.1 Modelo Auto-Regressivo (AR)	17

3.2.1.2 Modelo de Médias Móveis (MA)	18
3.2.1.3 Modelo Auto-Regressivo e de Médias Móveis (ARMA)	19
3.2.2 Modelos não estacionários	19
3.2.2.1 Modelos Auto-Regressivos Integrados de Médias Móveis (ARIMA)	19
3.3 ETAPAS DA METODOLOGIA DE BOX & JENKINS	21
3.3.1 Identificação	21
3.3.1.1 Identificação da ordem de integração	22
3.3.1.2 Determinação de p e q	23
3.3.2 Estimação	23
3.3.3 Verificação	23
3.3.4 Previsão	24
3.4 MODELOS SAZONAIS	25
3.4.1 Modelo Sazonal Auto-Regressivo de ordem P - $SAR(P)_s$	25
3.4.2 Modelo Sazonal de Médias Móveis de ordem Q - $SMA(Q)_s$	26
3.4.3 Modelo Sazonal Auto-Regressivo de Médias Móveis - $SARMA(P, Q)_s$	26
3.4.4 Modelo Sazonal Auto-Regressivo Integrado de Médias Móveis $SARIMA(P, D, Q)_s$	26
3.4.5 Modelo sazonal multiplicativo geral	26
3.4.6 Identificação	27
3.5 CONSIDERAÇÕES FINAIS	27
CAPÍTULO 4 - REDES NEURAIS ARTIFICIAIS PARA PROCESSAMENTO TEMPORAL	29
4.1 INTRODUÇÃO	29
4.2 TAXINOMIA DE MOZER	33
4.2.1 Formas de memória de curto-tempo	35
4.2.1.1 Memória Tapped Delay-line	35
4.2.1.2 Memória Exponencial	37
4.2.1.3 Memória Gama	37
4.2.1.4 Outras formas de memórias	38
4.2.2 Conteúdo da memória de curto-tempo	38
4.2.3 Adaptabilidade da memória	42
4.3 ARQUITETURA PERCEPTRON MÚLTIPLAS CAMADAS - MLP	44
4.4 ARQUITETURA DE REDES NEURAIS RECORRENTES PARA PREVISÃO DE SÉRIES TEMPORAIS	47
4.5 APLICAÇÕES DE RNAS À PREVISÃO	48
4.6 CONSIDERAÇÕES FINAIS	53

CAPÍTULO 5 - MODELAGEM DE REDES NEURAIS ARTIFICIAIS PARA PREVISÃO DE SÉRIES DE TEMPO	54
5.1 CONSIDERAÇÕES NO PROJETO DE RNAS	54
<i>5.1.1 Fatores determinantes no desempenho das RNAs</i>	<i>54</i>
<i>5.1.2 Controle dos fatores determinantes no desempenho das RNAs</i>	<i>56</i>
<i>5.1.2.1 Padronização</i>	<i>56</i>
<i>5.1.2.2 Função de saída ou transferência</i>	<i>59</i>
<i>5.1.2.3 Função de erro</i>	<i>60</i>
<i>5.1.2.4 Parâmetros de controle do gradiente descendente e do tempo de treinamento</i>	<i>61</i>
<i>5.1.2.5 Arquitetura da RNA</i>	<i>62</i>
5.2 CONSIDERAÇÕES FINAIS	62
CAPÍTULO 6 - ESTUDO DE CASO	64
6.1 DADOS	64
6.2 CARACTERÍSTICAS NÃO-OBSERVÁVEIS DA SÉRIE	67
6.3 MODELOS PARA PREVISÃO DE PREÇO DE OVOS	70
6.3.1 RNA	70
6.3.2 Modelo SARIMA	72
6.4 AVALIAÇÃO DA RNA	75
6.5 RESULTADOS OBTIDOS	75
CAPÍTULO 7 - ANÁLISE E DISCUSSÕES	78
7.1 REPRESENTAÇÃO	78
7.2 ROBUSTEZ DO MODELOS DE RNA	80
7.3 GENERALIZAÇÃO	81
7.4 APRENDIZADO EM RNAS	82
CAPÍTULO 8 - CONCLUSÕES FINAIS	84
CAPÍTULO 9 - TRABALHOS FUTUROS	87
REFERÊNCIAS BIBLIOGRÁFICAS	88
ANEXO A	95
ANEXO B	101

LISTA DE FIGURAS

Figura 2.1 - Preço real do ovo branco (tipo grande) pago ao produtor de São Paulo: Jan/85 a Dez/99	08
Figura 3.1 - Diagrama operacional do modelo Box & Jenkins.....	27
Figura 4.1 - Formulação abstrata da tarefa de predição temporal.....	34
Figura 4.2 - Função núcleo para memórias Delay-line, Exponencial, Gama e Gaussiana.....	36
Figura 4.3 - Modelo de memória Tapped Delay-line, segundo a taxinomia de Mozer.....	37
Figura 4.4 - Arquitetura de rede neural padrão e memória TIS.....	40
Figura 4.5 - Arquitetura da memória I e TI.....	41
Figura 4.6 - Arquitetura da memória TIS.....	41
Figura 4.7 - Arquitetura da memória O e TO.....	41
Figura 4.8 - Arquitetura da memória TOS.....	41
Figura 4.9 - Topologia genérica do Perceptron Multicamadas.....	45
Figura 6.1 - Preço real do ovo branco (tipo grande) pagos ao produtor de São Paulo: Jan/85 a Dez/99.....	65
Figura 6.2 - Preço real do ovo branco (tipo grande) pagos ao produtor de São Paulo, após a transformação.....	66
Figura 6.3 - Conjunto de treinamento utilizado. Dados de Jan/85 a Dez/97.....	66
Figura 6.4 - Conjunto de teste utilizado. Dados de jan/98 a Dez/99.....	67
Figura 6.5 - Componente sazonal da série.....	68
Figura 6.6 - Componente tendência/ciclo da série.....	69
Figura 6.7 - Componente irregular da série.....	70
Figura 6.8 - Arquitetura de RNA proposta para a previsão de preço de ovos.....	71
Figura 6.9 - Resíduos do modelo SARIMA (5,1,0)(0,0,2) ₁₂	72
Figura 6.10 - Função de autocorrelação (FAC) dos erros do modelo SARIMA (5,1,0)(0,0,2) ₁₂	73
Figura 6.11 - Função de autocorrelação parcial (FACP) dos erros do modelo SARIMA (5,1,0)(0,0,2) ₁₂	74
Figura 6.12 - Gráfico normal dos resíduos - modelo SARIMA (5,1,0)(0,0,2) ₁₂	74
Figura 6.13 - Comparação dos resultados obtidos.....	76
Figura 6.14 - Gráfico de dispersão dos erros obtidos, após a aplicação dos modelos.....	76
Figura A.1 - Histograma da variável em estudo (Preço do ovo branco (tipo grande) pagos ao produtor de São Paulo): Jan/85 a Dez/99.....	95
Figura A.2 - Representação da distribuição do preço do ovo em diagramas de caixas.....	95

Figura A.3 - Representação da distribuição mensal do preço do ovo em diagramas de caixas.....	96
Figura A.4 - Função de Autocorrelação (FAC) da série transformada.....	96
Figura A.5 - Função de Autocorrelação Parcial (FACP) da série transformada.....	97
Figura A.6 - Função de Autocorrelação (FAC) da primeira diferença sazonal da série transformada.....	97
Figura A.7 - Função de Autocorrelação Parcial (FACP) da primeira diferença sazonal da série transformada.....	98
Figura A.8 - Função de Autocorrelação (FAC) da série após duas diferenciações.	98
Figura A.9 - Função de Autocorrelação Parcial (FACP) da série após duas diferenciações.....	99
Figura A.10 - Resíduos da série diferenciada.....	100

LISTA DE TABELAS

Tabela 6.1 - Parâmetros estimados para o modelo $(0, 1, 0)(2, 1, 0)_{12}$	73
Tabela 6.2 - Comparação dos resultados obtidos.....	75
Tabela 6.3 - Medidas descritivas dos erros cometidos pelos dois modelos no teste de previsão (1988 e 1999).....	77
Tabela A.1 - Medidas descritivas da variável em estudo.....	100
Tabela B.1 - Preço real do ovo branco (tipo grande) pagos ao produtor de São Paulo: Jan/85 a Dez/99.....	101
Tabela B.2 - Conjunto de treinamento da rede.....	103

LISTA DE QUADROS

Quadro 2.1 - Principais medidas de desempenho.....	13
Quadro 3.1 - Comportamento da FAC e da FACP característico dos processos AR, MA, ARMA.....	23
Quadro 4.1 - Taxinomia de Horne & Giles.....	31
Quadro 4.2 - Taxinomia de Tsoi & Back.....	32
Quadro 4.3 - Taxinomia de Mozer.....	33
Quadro 4.4 - Resumo dos principais tipos de transformação e formas de memória	42
Quadro 4.5 - Arquitetura de redes neurais para processamento temporal.....	44
Quadro 4.6 - Resumo das principais aplicações de RNAs na previsão de séries de tempo.....	50
Quadro 5.1 - Exemplos de padronização não-linear.....	58
Quadro 5.2 - Principais fatores determinantes no desempenho das RNAs.....	63

LISTA DE REDUÇÕES

APA	Associação Paulista de Avicultura
APOESP	Associação de Avicultores do Oeste do Estado
BPTT	Backpropagation Through Time
FAC	Função de Autocorrelação
FACI	Função de Autocorrelação Inversa
FACP	Função de Autocorrelação Parcial
FACPI	Função de Autocorrelação Parcial Inversa
FIR	Finite Duration Impulse Response
GMRAE	Média Geométrica do Erro Absoluto Relativo
IEA	Instituto de Economia Agrícola
MAD	Desvio Absoluto Médio
MAE	Erro Absoluto Médio
MAPE	Erro Percentual Médio Absoluto
MdAPE	Erro Percentual Absoluto Médio
ME	Erro Médio
MLP	Perceptron Múltiplas Camadas
MPE	Erro Percentual Médio
MSE	Erro Quadrático Médio
NRMSE	Erro Quadrático Médio Normal Absoluto
PE	Erro Percentual
RAE	Erro Absoluto Relativo
RBF	Radial Basis Function
RMSE	Erro Quadrado Médio Absoluto
RNA	Rede Neural Artificial
RTRL	Real-Time Recurrent Learning
SARIMA	Modelo Sazonal Auto-Regressivo Integrado de Médias Móveis
TDNN	Time Delay Neural Network
U_2	Estatística U_2 de Theil

RESUMO

TÁPIA, Milena. **Redes Neurais Artificiais: Uma Aplicação na Previsão de Preços de Ovos**. Florianópolis, 2000. 105f. Dissertação (Mestrado em Ciências da Computação) - Curso de Pós-Graduação em Ciências da Computação, Universidade Federal de Santa Catarina.

Pesquisa que aborda o uso de Redes Neurais Artificiais (RNAs) - modelos biologicamente inspirados - em problemas de processamento temporal, onde o principal objetivo é a previsão. A previsão da série temporal utilizada neste trabalho, preço do ovo branco (tipo grande) pago ao produtor, serve como fonte de informação, para o produtor avícola, nas decisões quanto a organização, programação e comercialização da produção. O uso de uma RNA justifica-se devido a dificuldade dos métodos estatísticos tradicionais em conseguir previsões acuradas a longo prazo, de séries temporais econômicas. Com base na Taxinomia de MOZER (1994) para processamento temporal, o foco do estudo recaiu em duas questões: 1) Definir a forma de memória de curto tempo, o conteúdo que deveria ser armazenado nesta, e como seus parametros serão atualizados; 2) definir a topologia a rede (tamanho, estrutura e conexões), assim como os parâmetros do algoritmo de treinamento (taxa de aprendizado, termo de momento e outros). O modelo resultante foi comparado com a Metodologia de Box & Jenkins para modelos univariados, avaliado e criticado em termos de: capacidade representativa, processo de identificação e capacidade preditiva. Os resultados mostram que uma RNA, quando bem modelada, têm potencial para representar qualquer mapeamento complexo, não-linear, que pode governar mudanças em uma série de tempo. No estudo de caso foi possível prever o preço do ovo para um período de quatorze meses à frente.

Palavras-chave: Análise e Previsão de Séries Temporais, Redes Neurais Artificiais, Metodologia Box & Jenkins, Preço de Ovos Pagos ao Produtor.

ABSTRACT

TÁPIA, Milena. **Redes Neurais Artificiais: Uma Aplicação na Previsão de Preços de Ovos**. Florianópolis, 2000. 105f. Dissertação (Mestrado em Ciências da Computação) - Curso de Pós-Graduação em Ciências da Computação, Universidade Federal de Santa Catarina.

This research approaches the use of Artificial Neural Networks (ANNs) - models biologically inspired - in problems of temporal sequence processing, where the main objective is forecast. The forecast of the time series used in this work, the producer price of white egg, is good as source of information for the producer in the process of decision make like the organization, programming and commercialization of the production. The use of a ANN is justified due to the difficulty of the traditional statistical methods in getting acurated long term forecast, of economic time series. Based in MOZER (1994) Taxonomy for time-varing patterns, the focus of the study relapsed in two subjects: 1) to define the form of the short-term memory, the content that would be stored in this, like as how the parameters in the model will be adjusted; 2) to define the topology of the net (size, structure and connections), as well as the parameters of the training algorithm (learning rate, term of moment and others). The resulting model was compared with the Box & Jenkins Methodology for univaried models, appraised and criticized in terms of: representative ability, the model building process and the applicability of the neural net approach. The results show that a RNA, when well modeled, have potential to represent any complex, no-linear mapping, that can govern changes in a time series. In the case study it was possible to foresee the price of the egg for a period of fourteen months ahead.

Key-works: Time Series Analysis and Forecast, Artificial Neural Networks, Box & Jenkins Methodology, Producer Price of White Egg.

CAPÍTULO 1 - INTRODUÇÃO

A observação da realidade é consequência da manifestação de eventos complexos e incertos no decorrer do tempo. Embora estes eventos não sejam exatamente iguais, eles também não são completamente diferentes. Há uma linha de continuidade, similaridade e predibilidade nestes eventos que permite prever eventos futuros, freqüentemente de modo correto, a partir de experiências passadas (MUELLER, 1996).

O desejo de prever o futuro e entender o passado levou à procura de leis que explicam o comportamento do fenômeno observado; exemplos variam desde a irregularidade de uma batida de coração, até a volatilidade do mercado econômico.

Se o conhecimento exato das leis que regem um determinado fenômeno é expresso através de equações subjacentes determinísticas conhecidas, em princípio, estas podem ser resolvidas para prever o comportamento futuro do sistema com base no conhecimento das condições iniciais. Entretanto, se as equações não são conhecidas, para se fazer uma previsão, é preciso encontrar o estado atual do sistema e ambas as regras que governam a sua evolução.

Existe uma grande classe de fenômenos (físicos, econômicos e outros) cujas equações subjacentes não são conhecidas; as regras que governam a evolução do sistema devem ser inferidas de regularidades no passado.

Para esta classe de fenômenos complexos que não podem ser expressos por um conjunto de fórmulas, as redes neurais artificiais provê um formalismo atraente. Em Redes Neurais Artificiais (RNAs), o procedimento usual na solução de problemas passa inicialmente por uma fase de aprendizagem, em que um conjunto de exemplos é apresentado para a rede, a qual extrai automaticamente as características necessárias para representar a informação fornecida.

Devido a sua natureza indutiva, as RNAs podem sobrepor a formulação teórica e inferir relações não-lineares complexas entre as observações de uma série temporal.

Uma RNA não depende de qualquer suposição relativa ao modelo subjacente. É adaptável e capaz de aprender inter-relações não-lineares dos dados, se estas existirem.

Uma vantagem adicional das RNAs é que é possível fazer uso extenso do conhecimento à priori que se tem sobre o problema, escolhendo exemplos de aprendizagem adequados.

Desta forma, as RNAs têm recebido considerável atenção nos últimos anos, sendo aplicadas em um grande número de situações e demonstrando desempenho superior às abordagens convencionais. Para problemas de processamento temporal, exemplos podem ser encontrados em FERNANDES et al. (1995), CORRÊA & PORTUGAL (1995), BRAGA et al. (1996), TAMARI et al. (1996), VALENÇA & VASCONCELOS (1997) e ZHANG et al. (2000), entre outros.

A previsão a longo prazo (um ano ou mais) de séries econômicas, como o preço de produtos agrícolas, requer métodos de previsão diferentes, uma vez que muitos fatos podem ocorrer, mudando substancialmente os padrões estabelecidos e/ou as relações existentes.

Assim sendo, este trabalho tem como objetivo principal discutir e avaliar a habilidade dos modelos conexionistas em realizar previsões acuradas, a longo prazo, de uma série temporal de preços de ovos pagos ao produtor.

Para avaliar o modelo, os resultados alcançados com a RNA serão comparados com os obtidos pela aplicação do modelo SARIMA, modelo estatístico univariado de previsão de séries temporais mais utilizado no tratamento de séries econômicas.

Aspectos como a qualidade das previsões, habilidade de representação, facilidade de utilização e aplicabilidade do modelo também serão discutidos.

1.1 NATUREZA DO PROBLEMA

A indústria avícola, como qualquer outra atividade agrícola, é caracterizada pelo alto grau de incerteza associado a seus processos decisórios.

No processo de comercialização de ovos, por exemplo, o tipo e o tamanho dos ovos devem ser selecionados de acordo com as exigências do mercado. Além disso, os custos de produção e os preços do mercado apresentam grandes flutuações durante o ano, dificultando na organização, programação e comercialização da produção.

Além do investimento extra que se tem na formação das poedeiras, cuja formação demora quase cinco meses, deve-se ter em mente que o produto é perecível e a produção se dá em fluxo constante, não podendo ser ajustada rapidamente às eventuais tendências de saturação do mercado, sob pena de queima de capital (MARTINS, 1996).

Como o ovo é consumido, predominantemente, in natura e a ração é geralmente feita pelos próprios criadores, nesta atividade, é o próprio criador que administra e toma as decisões de ampliação ou redução da produção (MARTINS, 1999). Não existe diferenciação de produto, pois a competição entre os produtores se dá em torno de preços e não em torno de marcas. O sistema de preços atua, então, como um sinalizador das principais tendências dos produtores e consumidores.

Da maneira como está organizada, a avicultura de postura praticamente se resume à estratégia de alojar mais ou menos pintos e antecipar ou não o descarte das poedeiras, sendo que este é um prejuízo razoável e um risco de perda de lucro, caso os preços venham a se recompor (MARTINS, 1996).

Porém, apesar da complexidade da atividade, essa, quando bem planejada, se caracteriza normalmente por uma rentabilidade constante e segura durante o ano inteiro.

Para se ter uma idéia, uma ave consome 6,5 kg de ração durante dos 4,43 meses de formação e 40,15 kg durante um ano de produção. A vida total é de 26,8 meses, sendo que a fase produtiva dura 22,4 meses, 83,58% da vida total. Põe 22,08 dúzias de ovos por ano de produção, que significa uma produtividade de 72,6 % ovos/dia/poedeira do plantel ou postura em 72,6% dos dias do ano por ave (MARTINS, 1996).

Segundo HUNTON (1999), as aves possuem capacidade genética para atingir de 87 a 98% da produção de ovos no pico de postura (6,17 meses após o início da produção).

Desse modo, prevendo o preço futuro de venda do ovo, o produtor pode organizar a produção de modo que o pico de postura coincida com o mês em que o preço está em alta, maximizando assim, o seu lucro e diminuindo seus riscos. Além disso,

dependendo do preço, o produtor poderá ou não levar o lote de aves a um segundo ciclo de postura (Muda Forçada) ou antecipar o descarte das poedeiras¹.

Hoje, as fontes de informação sobre preço de ovos a nível de produtor são o Instituto de Economia Agrícola - IEA, que divulga preços diários na granja em várias regiões do estado, e a Associação de Avicultores do Oeste do Estado - APOESP, cujos preços referem-se ao produto posto em São Paulo, Capital.

Uma previsão para um período maior (30 dias) é fornecida pela Associação Paulista de Avicultura - APA, em sua revista *Aves & Ovos* publicada mensalmente. Porém, essa informação é insuficiente para predizer a rentabilidade do investimento. Para isso, uma previsão para um período maior de tempo é necessária.

Neste trabalho optou-se pelo preço de ovos brancos (tipo grande) pagos ao produtor pelos seguintes motivos:

- Segundo um estudo feito por MARTINS (1996), os ganhos de produtividade na cadeia produtiva de ovos, - obtidos pela inclusão de novas tecnologias, melhoramento genético das aves, entre outros fatores - são captados, em grande parte, pelos atacadistas e supermercados. No período analisado pela autora (1985 à 1995), o resultado econômico ou margem bruta do criador foi negativo. Assim sendo, o risco que o criador corre em relação ao preço, é 50% maior do que para os atacadistas e varejistas.
Incluindo a estimativa do preço futuro como informação no processo de tomada de decisão do agricultor, espera-se diminuir os riscos e o prejuízo.
- Ainda segundo a autora, a produção de ovos brancos (tipo grande) é maior. Porém, o comportamento dos diversos tipos de ovos (extra, grande, médio e pequeno), como também nos vários mercados (preço ao produtor, no atacado e varejo) é quase constante, ou seja, sobem ou caem juntos, na mesma proporção.

¹ *Muda Forçada*: Consiste em levar um lote em postura a uma situação crítica de *stress* alimentar, em que as aves perderão cerca de 25 % de seu peso vivo, ocorrendo uma muda drástica de penas. Depois da muda forçada, as aves voltam à produção com postura bem acima do nível anterior, com um novo pique de produção, e permanecem em postura por mais um período de 30 semanas, aproximadamente (NORTH & BELL, 1990).

A escolha deste tema mostra também, como técnicas de Inteligência Artificial - um paradigma que até bem pouco tempo era considerado 'inútil' para resolução de problemas reais - podem ser aplicadas com sucesso aos mais diversos tipos de problemas, neste caso, a Avicultura de postura.

1.2 OBJETIVOS DA PESQUISA

1.2.1. OBJETIVO GERAL

Propor, validar e verificar um modelo de RNA para previsão de preço de ovos para um período de no mínimo 10,6 meses à frente (período da formação das aves ao pico da produção).

1.2.2. OBJETIVOS ESPECÍFICOS

- Rever os principais conceitos e modelos de Análise e Previsão de Séries Temporais;
- avaliar os diversos tipos de RNAs disponíveis, objetivando resolver o problema de processamento temporal;
- estudar o comportamento da série de preço de ovos a procura de periodicidades relevantes nos dados, que poderão ser utilizadas na modelagem da RNA;
- propor um modelo de RNA para o problema em questão;
- comparar o desempenho do modelo proposto, em termos de: capacidade representativa, processo de identificação e capacidade preditiva, com o desempenho dos métodos de previsão tradicionais, em especial, a metodologia de Box & Jenkins para modelos univariados;
- e levantar os aspectos positivos e negativos do modelo para problemas de processamento temporal.

1.3 ORGANIZAÇÃO DO TRABALHO

O restante desta dissertação está estruturada da seguinte forma:

Do Capítulo 2 ao Capítulo 5, é apresentada a fundamentação teórica adotada para tratar o tema e o problema de pesquisa.

Na primeira parte, Capítulo 2, é feito um estudo sobre séries temporais, onde os pontos enfatizados são: os objetivos na análise de séries temporais e os principais métodos estatísticos para previsão. Um levantamento das principais medidas utilizadas para avaliar o desempenho dos métodos de previsão fecha este capítulo.

O Capítulo 3 descreve a metodologia Box & Jenkins para modelos univariados, que é o método estatístico mais utilizado para previsão de séries econômicas e utilizado neste trabalho para avaliar o desempenho do modelo proposto. São descritos alguns modelos, assim como o processo de identificação, estimação, verificação e previsão.

No Capítulo 4, é estudado o uso de Redes Neurais Artificiais - RNAs para processamento temporal. A taxinomia de MOZER (1994) é abordada, assim como, as redes neurais diretas. Conclui-se esta parte com uma rápida revisão dos principais trabalhos de processamento temporal envolvendo RNAs.

O Capítulo 5 enfoca todo o processo de modelagem da RNA para previsão de séries temporais. Os mecanismos de controle de desempenha que são considerados durante a modelagem, tais como: a escolha da função de ativação, a escolha da função erro, a escolha dos parâmetros de controle e da arquitetura da rede, são analisados separadamente, destacando-se a influência de cada um no que diz respeito à convergência, generalização e escalabilidade da rede.

O Capítulo 6 descreve o estudo de caso realizado. Primeiramente, é feito um estudo minusioso da série de preço de ovos. Após, são descritas todas as etapas da modelagem da RNA e da identificação do modelo SARIMA para previsão de preço de ovos.

O Capítulo 7 discute a habilidade de representação, o processo de cosntrução e a aplicabilidade das RNAs em comparação com a metodologia Box & Jenkins.

E para finalizar, os resultados obtidos são interpretados e analisados (Capítulo 8).

CAPÍTULO 2 - MÉTODOS ESTATÍSTICOS DE PREVISÃO

Definir o que é série temporal, análise de séries temporais e seus objetivos; mostrar as formas clássicas de classificação dos métodos de previsão de séries temporais e as principais medidas utilizadas na avaliação de desempenho.

Os métodos estatísticos de previsão variam muito, podendo ser simples e intuitivos com pouca ou nenhuma análise dos dados, ou mais quantitativos e complexos, com uma análise considerável de dados.

MAKRIDAKS et al. (1998) classifica os métodos estatísticos de previsão em duas categorias: Quantitativos e Qualitativos.

Os métodos quantitativos se subdividem em Séries Temporais e Modelos Econométricos. Ambos são projeções técnicas que envolvem ajustes dos modelos teóricos para os dados a serem analisados. A suposição utilizada é que as observações da amostra de um ou mais processos, aparentemente aleatórios, fornecem evidências confiáveis sobre o comportamento futuro do processo em estudo (MAKRIDAKIS et al. 1998). A diferença entre eles é que o procedimento econométrico se baseia fortemente na teoria econômica para construir o modelo, a partir de uma ou mais relações causais entre variáveis. Já no procedimento de séries temporais, em princípio, apenas os dados da própria série são utilizados para construir o modelo. Quando outras variáveis explicativas são incorporadas ao modelo, este não tem que se harmonizar necessariamente, com a teoria econômica. O importante é que o modelo produza bons resultados.

Os métodos de previsão qualitativos, por sua vez, foram projetados para prever o comportamento futuro de processos onde dados históricos não existem ou não estão disponíveis, mas, têm-se conhecimento qualitativo disponível suficiente. Esses modelos confiam na análise de entradas subjetivas obtidas de várias fontes, tais como: opiniões de especialistas, executivos, carências de consumo, entre outras.

2.1 ANÁLISE DE SÉRIES TEMPORAIS

2.1.1. CONCEITOS BÁSICOS

De uma forma geral, pode-se definir uma série temporal como qualquer conjunto de observações ordenadas seqüencialmente no tempo ABELÉM (1994), MAKRIDAKIS et al. (1998), FAVA (2000a) (Fig. 2.1). Uma enorme quantidade de fenômenos de natureza física, biológica, econômica, e outras, estão enquadrados nesta categoria e têm suas características estudadas através da análise de séries temporais (FAVA, 2000a).

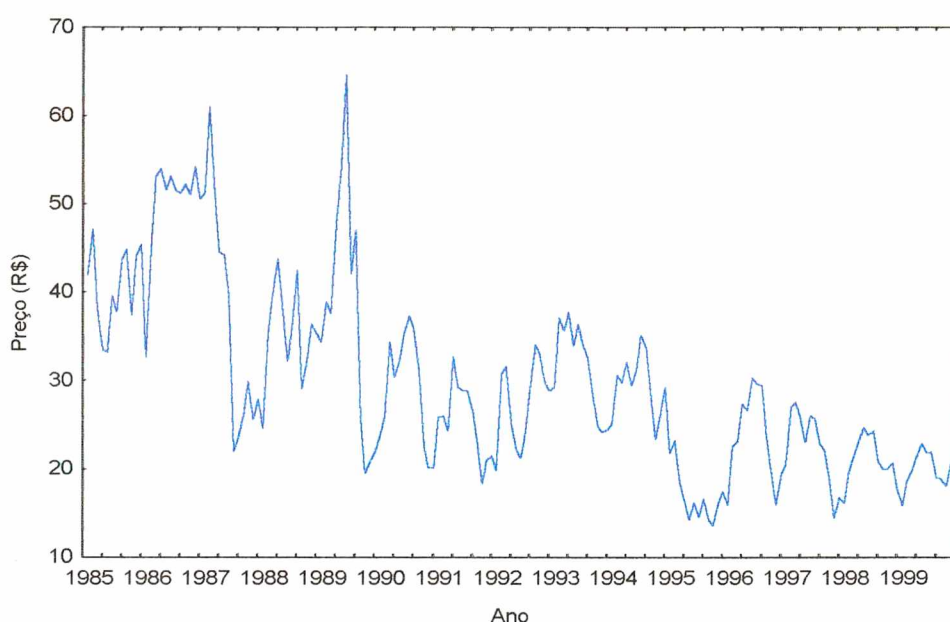


Figura 2.1 - Preço real do ovo branco (tipo grande) pago ao produtor de São Paulo: Jan/85 a Dez/99.

Segundo MILONE & ANGELINI (1995), o gráfico de uma série temporal pode ser descrito por um ponto que se move com o decorrer do tempo, de alguma forma análogo à trajetória de uma partícula material que se desloca sob a influência de forças físicas. Entretanto, o movimento pode ser provocado, em vez de forças físicas, por uma combinação de forças econômicas, sociológicas, psicológicas e outras.

Esses movimentos característicos das séries temporais podem ser classificados em quatro tipos principais, freqüentemente denominados componentes de uma série temporal:

- **Tendência (T):** movimento contínuo dos dados numa determinada direção, que dura por um período de tempo grande;
- **Ciclo (C):** movimento oscilatório em torno da tendência, que se verifica durante um período de tempo maior que um ano e que não é necessariamente regular;
- **Sazonalidade (S):** movimento oscilatório que ocorre com regularidade em subperíodos de um período de tempo fixo, geralmente um ano;
- **Componente Errático (E):** movimento esporádico, irregular, que decorre de fatores estritamente aleatórios.

Sendo assim, uma série temporal pode resultar da combinação de todos os quatro componentes mencionados ou de um subconjunto deles. Tratando-se de séries econômicas, de acordo com FAVA (2000a), o termo errático estará sempre presente.

2.1.2. OBJETIVOS

Os principais objetivos na análise de séries temporais é:

- entender o mecanismo do sistema gerador da série temporal e,
- prever o comportamento futuro do sistema.

Para isso, os problemas de interesse são:

- descrever o comportamento da série;
- investigar o mecanismo gerador da série temporal;
- procurar periodicidades relevantes nos dados.
- fazer previsões de valores futuros da série, a partir de valores passados;

Há dois aspectos principais a serem considerados no estudo de séries temporais: *análise e modelagem*.

A principal razão para se modelar uma série temporal é tornar possível previsões de valores futuros, bem como geração de seqüências sintéticas equiprováveis. Como em

qualquer análise de dados, procura-se obter um modelo que descreva o sistema de maneira parcimoniosa¹ para o objetivo desejado.

O objetivo da análise da séries temporais é, então, sumarizar as propriedades da série e caracterizar seu comportamento, identificando ou sugerindo um modelo adequado.

Há, basicamente, dois enfoques usados nesta análise.

- No primeiro, a análise é feita no domínio de tempo, isto é, o interesse reside na magnitude de eventos que ocorrem em determinado instante de tempo e na relação entre observações em diferentes instantes de tempo. A ferramenta mais utilizada é a função de autocorrelação (e certas funções da mesma) e a análise é baseada em um modelo paramétrico²;
- No segundo, a análise é feita no domínio de frequência, ou seja, o interesse está na frequência com que certos eventos ocorrem em determinado período de tempo (movimentos cíclicos, por exemplo). A ferramenta utilizada é o espectro (que é uma transformação de Fourier da função de autocorrelação) e a análise é baseada em modelos não paramétricos³.

As duas formas de análise não são alternativas, mas sim complementares, mostrando cada uma diferentes aspectos da natureza da série temporal. Entretanto, pode-se dizer que a análise no domínio da frequência é conveniente na procura de características determinísticas da série temporal, enquanto a do domínio do tempo é usada na análise de processos não determinísticos.

2.1.3. PREVISÃO DE SÉRIES TEMPORAIS

Previsão pode ser definida como uma manifestação relativa a sucessos desconhecidos em um futuro determinado (MULLER, 1996). A previsão em si, não constitui um fim, mas um meio de fornecer informações e subsídios para uma

¹ Princípio que estabelece que deve-se escolher um modelo com o menor número de parâmetros possíveis, para uma adequada representação matemática (SOUZA & CAMARGO, 1996).

² Modelo que especifica certas condições (tais como: obs. Independentes, populações normais e homocedásticas, e outras) sobre os parâmetros da população da qual se extraiu a amostra.

conseqüente tomada de decisão, visando atingir determinado objetivo SOUZA & CAMARGO (1996) e MAKRIDAKIS et al. (1998).

Dependendo do valor assumido pelo horizonte de previsão, SOUZA & CAMARGO (1996) classificam as previsões dos valores futuros de uma série temporal como de prazo imediato; curto, médio ou longo prazo.

- **Prazo imediato:** menor que um mês (dados diários ou semanais);
- **Curto prazo:** de um a três meses (dados semanais ou mensais);
- **Médio prazo:** de três a seis meses (dados mensais ou trimestrais);
- **Longo prazo:** um ano ou mais (dados mensais, trimestrais ou anuais).

Desse modo, diante da possibilidade da existência de diferentes horizontes de previsão, MULLER (1996) especifica técnicas distintas para prognosticar os valores futuros de uma série temporal:

- Previsões "*multi-step*"
- Previsões "*single-step*"

Nas previsões "*multi-step*", o conjunto de valores correntes é empregado na realização da previsão para determinado instante; esta previsão é, então, introduzida entre as observações passadas, compondo um novo conjunto de dados, sobre o qual será obtida a previsão para o tempo subsequente.

Nas previsões "*single-step*" não há incorporação de previsões aos dados utilizados para encontrar a próxima previsão, sendo esta, independente dos valores anteriormente previstos. A previsão é feita apenas para o período de tempo imediatamente posterior ao atual, a partir das observações da série temporal.

Ainda, SOUZA (1989), de acordo com o número de séries temporais envolvidas na modelagem, classifica os métodos de previsão de séries temporais estatísticos em:

³ Modelo que não especifica condições sobre os parâmetros da população da qual se extraiu a amostra.

- **Modelos Univariados:** Grupo que inclui todos os métodos que se baseiam em uma única série histórica para a realização dos prognósticos;
- **Modelos de Função de Transferência:** Inclui todos os métodos que associam mais de uma série, com garantia de que a relação de causalidade entre estas séries é conhecida.
- **Modelos Multivariados:** Métodos que incluem simultaneamente duas ou mais séries temporais sem qualquer imposição com relação à causalidade entre essas séries.

Segundo SOUZA (1989), entre os modelos univariados, os que merecem destaque são: *decomposição por componentes não observáveis*, que foi o modelo mais utilizado até a década de 60; os *modelos automáticos ou caixa preta*, que surgiram na década de 60 e incluem os modelo de regressão, médias móveis, ajustamento sazonal e alisamento exponencial; *modelos de Box & Jenkins*, propostos no início da década de 70, consistindo de uma classe geral de modelos lineares conhecidos como ARIMA; *modelos bayesianos*, introduzidos também na década de 70 e que caracterizam-se pela possibilidade de incorporação de informações subjetivas no processo de modelagem da série.

Entre os modelos de função de transferência pode-se citar: modelos de função de transferência e de análise de intervenção, modelos estruturais e os modelos de função de transferência bayesiana.

Para a classe de modelos multivariados, podem ser destacados: *modelo MARMA* de Box & Tiao, *formulação do espaço de estado* e *formulação bayesiana multivariada*.

2.1.4. MEDIDAS DE DESEMPENHO

A suposição básica de qualquer técnica de previsão de séries temporais é que o valor observado na série fica determinado por um padrão que se repete no tempo e por alguma influência aleatória (MAKRIDAKIS et al, 1998). Ou seja, mesmo que o padrão exato que caracteriza o comportamento da série temporal seja isolado, algum desvio ainda existirá entre os valores futuros reais e os previstos. Essa aleatoriedade não pode

ser prevista, mas, se isolada, sua magnitude pode ser estimada e usada para determinar a variação ou erro entre as observações reais (Y_t) e as previsões realizadas (F_t).

As medidas mais utilizadas na avaliação dos métodos de previsão são citadas a seguir:

Quadro 2.1 - Principais medidas de desempenho

Erro Médio (ME)	$ME = \frac{1}{n} \sum_{t=1}^n (Y_t - F_t)$	(1)
Erro Absoluto Médio (MAE)	$MAE = \frac{1}{n} \sum_{t=1}^n (Y_t - F_t)$	(2)
Erro Quadrático Médio (MSE)	$MSE = \frac{1}{n} \sum_{t=1}^n (Y_t - F_t)^2$	(3)
Erro Quadrático Médio Normal Absoluto (NRMSE)	$NRMSE = \frac{\sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - F_t)^2}}{\sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \bar{Y})^2}}$	(4)
onde,		
$\bar{Y} = \frac{1}{n} \sum_{t=1}^n (Y_t)$		
Erro Percentual (PE_t)	$PE_t = \left(\frac{Y_t - F_t}{Y_t} \right) 100$	(5)
Erro Percentual Médio (MPE)	$MPE = \frac{1}{n} \sum_{t=1}^n PE_t$	(6)
Erro Percentual Médio Absoluto (MAPE)	$MAPE = \frac{1}{n} \sum_{t=1}^n PE_t $	(7)
Estatística U_2 de Theil (U_2)	$U_2 = \sqrt{\frac{\sum_{t=1}^{n-1} \left(\frac{F_{t+1} - Y_{t+1}}{Y_t} \right)^2}{\sum_{t=1}^{n-1} \left(\frac{Y_{t+1} - Y_t}{Y_t} \right)^2}}$	(8)

- se $U_2 = 0$, as variações previstas são iguais às observadas (previsão perfeita);
- se $U_2 \geq 1$, o erro de previsão do modelo é maior que o erro de uma previsão trivial, que diz que "a melhor estimativa para o preço de amanhã é o preço de hoje".
- quanto menor o valor de U_2 , melhor o modelo de previsão.

ARMSTRONG & COLLOPY (1992) analisaram várias medidas de erro disponíveis segundo a confiabilidade, validade, sensibilidade para pequenas mudanças,

proteção contra observações aberrantes (*outliers*) e facilidade de interpretação. Os estudos levaram as seguintes conclusões:

- Quando a tarefa envolve calibrar um modelo para um conjunto de séries temporais, a melhor medida é a GMRAE (Média Geométrica do Erro Absoluto Relativo);
- Quando o objetivo é selecionar o método de previsão mais preciso, as medidas indicadas são RAE (Erro Absoluto Relativo), Estatística U2 de Theil e MdAPE (Erro Percentual Absoluto Mediano).
- O RMSE (Erro Quadrático Médio Absoluto), muito utilizado na literatura, foi considerado impróprio para comparar a precisão dos métodos de previsão.

Dos trabalhos encontrados na literatura, onde a maioria tem por objetivo avaliar a capacidade preditiva de dois ou mais métodos, seis utilizaram a medida MAPE (19,35%), cinco utilizaram a medida MAE (16,13%), quatro utilizaram a medida MPE (12,90%) e quatro utilizaram a medida MSE (12,90%). Os outros 38,72% utilizaram outras medidas, como: NMSE, RMSE, PE e outras.

A medida MAPE, segundo ARMSTRONG & COLLOPY (1992), atribui uma penalidade muito maior aos valores previstos que excedem o valor desejado, do que aos valores previstos que são menores que o desejado. Sendo assim, a medida MAPE é desaconselhada. Uma alternativa, segundo AHLBURG (1992), seria o RMdAPE (Erro Percentual Absoluto Mediano Relativo), que corrige os problemas do MAPE.

Segundo TAYLOR (1992), para a avaliação de diferentes métodos para previsão de uma única série, o MSE poderia ser usado. Porém, quando se aplica o mesmo método para um grupo de séries, o uso do MSE pode levar a resultados desastrosos.

Uma solução, para o problema da escolha de uma medida de erro adequada, seria, como utilizado por MAKRIDAKIS et al.(1998), incluir na avaliação, a maioria das medidas padrões de erro.

2.2 CONSIDERAÇÕES FINAIS

A análise de séries temporais tem dois objetivos principais: prever o comportamento futuro do sistema e entender o mecanismo gerador. O objetivo da previsão é prever com precisão a evolução do sistema. Na previsão, não importa o quanto bem um modelo ajusta aos dados - só a qualidade das previsões futuras importa (WEIGEND & GERSENFLD, 1994).

Já para entender o mecanismo gerador do sistema, procura-se uma descrição, que com precisão, capture características do comportamento do sistema a longo prazo.

A suposição básica tanto para análise, como para a previsão é que o valor observado na série fica determinado por um padrão que se repete no tempo e por alguma influência aleatória.

Visando determinar o padrão que se repete no tempo, vários métodos foram criados. Estes variam de simples e intuitivos a quantitativos e complexos. A escolha de qual usar, depende do tipo de informação que se têm disponível (quantitativa ou qualitativa).

Dependendo do número de séries temporais envolvidas na previsão, pode-se classificar os métodos de previsão em univariados, função de transferência e multivariados.

A influência aleatória, não pode ser prevista, mas, se isolada, sua magnitude pode ser estimada e usada para medir o desempenho do método de previsão.

CAPÍTULO 3 - METODOLOGIA DE BOX & JENKINS PARA MODELOS UNIVARIADOS

*Descrever a metodologia de Box & Jenkins para modelos univariados,
identificar os modelos possíveis dentro desta e os passos para sua aplicação
na previsão de séries temporais;*

3.1 INTRODUÇÃO

A análise de séries temporais, segundo o enfoque de BOX & JENKINS (1976), tem como objetivo principal a realização de previsão (MAKRIDAKIS et al., 1998). Essa metodologia permite que valores futuros de uma série sejam previstos tomando por base apenas seus valores presentes e passados. Isso é feito explorando a correlação temporal que geralmente existe entre os valores exibidos pela série.

De acordo com FAVA (2000b), a realização temporal considerada pelo enfoque de Box & Jenkins é representada formalmente por um conjunto de processos estocásticos genericamente denominados modelos ARIMA: em cada instante de tempo t , existe uma família de valores que a série pode assumir, aos quais estão associadas probabilidades de ocorrência¹.

Em tese, é possível que exista, para cada instante de tempo, uma particular função de densidade de probabilidade. Isso implica dizer que cada variável aleatória $y(t), t = t_1, t_2, \dots$, pode ter média e variância específicas.

A tarefa consiste, então, em descobrir qual é o processo gerador da série que se está estudando, ou seja, qual o modelo que traduz o mecanismo de geração da série.

Os modelos ARIMA resultam da combinação de três componentes também denominados "filtros": o componente Auto-regressivo (AR), o filtro de Integração (I) e o componente de Médias Móveis (MA)².

¹ Segundo (FAVA, 2000), processo estocástico é uma sequência ordenada (no tempo) de variáveis aleatórias $\{y(t), t \in T\}$ definidas no mesmo espaço de probabilidade (S, \mathfrak{F}, P) .

² A abreviatura MA vem da expressão Moving Average.

Uma série de tempo pode conter os três filtros ou apenas um subconjunto deles, resultando daí várias alternativas de modelos passíveis de análise pela metodologia de Box & Jenkins, conforme será visto a seguir.

Uma restrição que tem que ser imposta ao processo estocástico é que este tem que ser estacionário. De acordo com FAVA (2000b), um processo estocástico $\{y(t), t \in T\}$ é fracamente estacionário se as condições a seguir forem satisfeitas para qualquer t :

$$E[y(t)] = \mu \quad (9)$$

$$V[y(t)] = E[y(t) - \mu]^2 = \sigma^2 \quad (10)$$

$$E[(y(t) - \mu)(y(t - k) - \mu)] = f(k), k = 1, 2, \dots \quad (11)$$

As duas primeiras condições (Equações 9 e 10) indicam que a média e a variância de $y(t)$ são invariantes no tempo. A terceira condição impõe que as autocovariâncias não dependem do tempo; elas variam apenas em relação à distância (k) que separa as duas observações consideradas.

A metodologia de Box & Jenkins permite trabalhar com um caso particular de séries não estacionárias, conforme será visto adiante.

3.2 MODELOS POSSÍVEIS DENTRO DA METODOLOGIA DE BOX & JENKINS

3.2.1. MODELOS ESTACIONÁRIOS

3.2.1.1. MODELO AUTO-REGRESSIVO (AR)

De acordo com esse modelo, y_t é descrito apenas por seus valores passados e pelo ruído branco³ ε_t . O modelo auto-regressivo genérico, representado por $AR(p)$

³ Na definição de (MAKRIDAKIS et al., 1998), uma variável aleatória com média zero, variância constante e não autocorrelacionada é denominada ruído branco.

pressupõe que seja o resultado da soma ponderada de seus p valores passados y_t , além do ruído branco ε_t :

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (12)$$

Utilizando o operador de defasagem⁴ B , o modelo pode ser rescrito como:

$$y_t = \phi_1 B y_t + \phi_2 B^2 y_t + \dots + \phi_p B^p y_t + \varepsilon_t \quad (13)$$

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) y_t = \varepsilon_t \quad (14)$$

$$\phi(B) y_t = \varepsilon_t \quad (15)$$

Na engenharia esse filtro é chamado de IIR (*Infinite Impulse Response*), porque a saída do sistema continua após cessar a entrada (WEIGEND & GERSHENFLD, 1994).

3.2.1.2. MODELO DE MÉDIAS MÓVEIS (MA)

Por esse modelo, a série y_t resulta da combinação linear dos choques aleatórios (ruídos brancos) ocorridos no período corrente e nos períodos passados.

O modelo de médias móveis genérico envolve q valores defasados de ε e é indicado por $MA(q)$ cuja equação é:

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (16)$$

A série y_t resulta, portanto, da soma ponderada do valor corrente e dos q valores passados de ε . Como não se impõe a restrição de que a soma dos "pesos" do $q + 1$ ruídos brancos seja igual à unidade, a expressão "média móvel" não traduz com precisão a natureza do processo gerador de y_t (FAVA, 2000b).

Fazendo uso do operador de defasagem, o modelo $MA(q)$ passa a ser representado por:

$$y_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t \quad (17)$$

ou

⁴ $B^k y_t = y_{t-k}$ onde, k é igual ao período de tempo para trás.

$$y_t = \theta(B)\varepsilon_t \quad (18)$$

O modelo MA é também chamado de filtro FIR (*Finite Impulse Response*), onde a saída do sistema é zero N passos de tempo depois que a entrada do sistema se torna zero (WEIGEND & GERSHENFLD, 1994).

3.2.1.3. MODELO AUTO-REGRESSIVO DE MÉDIAS MÓVEIS (ARMA)

Conforme indica o próprio nome, esse modelo é uma combinação dos dois anteriores: y_t é descrito por seus valores passados e pelos choques aleatórios corrente e passado.

A especificação genérica de um modelo ARMA admite um componente auto-regressivo de ordem p e um componente de médias móveis de ordem q : ARMA(p, q), cuja equação é:

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (19)$$

Usando o operador de defasagem B , resulta:

$$(1 - \phi_1 B - \dots - \phi_p B^p) y_t = (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t \quad (20)$$

$$\phi(B)y_t = \theta(B)\varepsilon_t \quad (21)$$

3.2.2. MODELOS NÃO ESTACIONÁRIOS

3.2.2.1. MODELOS AUTO-REGRESSIVOS INTEGRADOS DE MÉDIAS MÓVEIS (ARIMA)

A metodologia de Box & Jenkins aplica-se a um caso específico de séries não estacionárias - séries que se tornam estacionárias após a aplicação de diferenças. O número necessário de diferenças para tornar uma série estacionária é *denominado ordem de integração (d)*.

Um exemplo pode ser encontrado em (FAVA, 2000b; p. 210). Conforme o autor, "se y_t é não estacionária mas $x_t = \Delta y_t = y_t - y_{t-1}$ é estacionária, então y_t é dita

integrada de ordem 1⁵. Se y_t precisar de duas diferenças para ser estacionarizada, ou seja, se $z_t = \Delta^2 y_t = \Delta(\Delta y_t) = \Delta(y_t - y_{t-1})$ é menor que estacionária, então y_t é integrada de ordem 2".

Uma série que exhibe tal comportamento é denominada "*não estacionária homogênea*". Ela apresenta tendência "*estocástica*", em contraposição à tendência determinística, que é expressa como função do tempo (FAVA, 2000b).

O modelo aplicado as séries não estacionárias homogêneas pode ser genericamente formulado da seguinte maneira: se y_t tornar-se estacionário após a aplicação de d diferenças e a série resultante for representada por um modelo ARMA (p, q), diz-se que y_t é descrita por um modelo ARIMA(p, d, q) representado por:

$$w_t = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (22)$$

$$\text{onde } w_t = \Delta^d y_t. \quad (23)$$

Alternativamente, utilizando o operador de defasagem, tem-se:

$$(1 - \phi_1 B - \dots - \phi_p B^p) w_t = (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t \quad (24)$$

$$w_t = (1 - B)^d y_t \quad \text{ou} \quad (25)$$

$$(1 - B)^d \phi(B) y_t = \theta(B) \varepsilon_t \quad (26)$$

Existem, também, os modelos ARI(p, d) ou ARIMA($p, d, 0$) e IMA(d, p) ou ARIMA($0, d, q$) que são as versões não estacionárias dos modelos AR e MA, respectivamente.

Um caso especial de processo não estacionário homogêneo é o ARIMA($0, 1, 0$), denominado *random walk* (ou passeio aleatório):

$$y_t = y_{t-1} + \varepsilon_t \rightarrow \Delta y_t = \varepsilon_t \quad (27)$$

A série y_t é integrada de *ordem 1* e sua primeira diferença é um ARMA($0, 0$), ou seja, um ruído branco. A denominação passeio aleatório decorre do fato de: a trajetória da série do instante $t - 1$ para t é inteiramente aleatória.

⁵ Δ é denominado operador diferença e guarda a seguinte relação com o operador defasagem: $\Delta = 1 - B$

3.3 ETAPAS DA METODOLOGIA DE BOX & JENKINS

Quando se aplica a metodologia de Box & Jenkins a uma série temporal, o primeiro passo consiste em descobrir qual entre as várias versões do modelo ARIMA descreve o comportamento da série. Essa etapa é denominada *identificação*. O segundo passo consiste em estimar os parâmetros do modelo identificado - os parâmetros ϕ se houver um componente auto-regressivo, os parâmetros θ se houver o filtro de médias móveis e a variância do ruído branco σ_ϵ^2 . Essa etapa é chamada *estimação*. Segue-se a ela a etapa de *verificação*, que consiste em avaliar se o processo de estimação foi bem-sucedido. Se a conclusão for negativa, repete-se o procedimento desde a etapa de identificação, com o objetivo de encontrar uma especificação de modelo mais adequada. Quando se obtém um modelo satisfatório, passa-se para a última etapa da metodologia, ou seja, a *previsão*, que se constitui no objetivo primordial do enfoque de Box & Jenkins.

Uma descrição sucinta de cada uma dessas etapas é apresentada a seguir.

3.3.1. IDENTIFICAÇÃO

O processo de identificação consiste em determinar quais dos filtros AR, I e MA compõem o processo gerador da série, bem como quais suas respectivas ordens.

De acordo com SOUZA & CAMARGO (1996), o processo de identificação dos modelos ARIMA consiste teoricamente de dois estágios 'distintos': 1) escolha adequada do grau de diferenciação " d " e 2) escolha adequada dos graus dos polinômios " p " e " q " do modelo ARMA aplicado à série w_t .

Em ambas as fases, a identificação da ordem do modelo (p, d, q) é feita através da Função de Autocorrelação - FAC e Função de Autocorrelação Parcial - FACP amostral.

A FAC amostral é constituída pela sequência de pares (k, r_k) , $k = 1, 2, \dots$, onde, r_k é o coeficiente de autocorrelação amostral de ordem k , e este é a ordem da correlação serial (*lag*), ou seja, a autocorrelação entre y_t e y_{t-k} .

A FACP amostral é constituída pela sequência de pares (k, ϕ_{kk}) , $k = 1, 2, \dots$, onde, ϕ_{jj} é o coeficiente de autocorrelação parcial de ordem kk , ou seja, a autocorrelação entre y_t e y_{t-k} depois que a influência de $y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$ sobre y_t foi descontada.

Uma descrição desses termos pode ser encontrada em (SOUZA & CAMARGO, 1996), (MAKRIDAKIS et al., 1998) e (FAVA, 2000b).

SOUZA & CAMARGO (1996) sugerem ainda, o uso da Função de Autocorrelação Inversa - FACI e da Função de Autocorrelação Parcial Inversa - FACPI na identificação do modelo.

3.3.1.1. IDENTIFICAÇÃO DA ORDEM DE INTEGRAÇÃO

O gráfico da série original pode dar uma primeira indicação da violação ou não da condição de estacionariedade (FAVA, 2000b).

Séries não estacionárias apresentam fortes correlações seriais, ou seja, os valores dos coeficientes de autocorrelação declinam muito lentamente à medida que k aumenta. Assim sendo, uma FAC amostral com valores inicialmente altos e que não declinam rapidamente para um valor estatisticamente igual a zero à medida que k aumenta, indica que a série é não estacionária e precisa, portanto, ser diferenciada (MAKRIDAKIS et al., 1998).

O número ótimo de diferenças é sempre igual a ordem de diferenciação cujo desvio padrão da série diferenciada é menor (INTRODUCTION TO ARIMA, 2000).

Segundo FAVA (2000b), nem sempre a aplicação de diferenças é suficiente para estacionarizar a série. Se esta exibir tendência determinística, é necessário removê-la antes de aplicar a metodologia de Box & Jenkins.

Se a não-estacionariedade estiver associada à variância da série, esta deve ser transformada antes de calcular sua diferença. As transformações mais utilizadas são o *logaritmo neperiano* (quando o desvio-padrão de y_t é proporcional ao seu nível) e a *raiz quadrada* (quando a variância de y_t é proporcional ao seu nível) (MAKRIDAKIS et al., 1998).

3.3.1.2. DETERMINAÇÃO DE P E Q

O próximo passo consiste em determinar se a série estacionária, original ou transformada, é gerada por um processo auto-regressivo ($p \neq 0$), de médias móveis ($q \neq 0$) ou misto ($p \neq 0$ e $q \neq 0$) e determinar o valor de p e/ou q .

As características da FAC e da FACP, resumidas no Quadro 3.1, é que indicarão qual o possível processo gerador da série.

Quadro 3.1 - Comportamento da FAC e da FACP característico dos processos AR, MA, ARMA.

Processo	FAC	FACP
AR(p)	Declinante	Nula para lags acima de $k = p$
MA(q)	Nula para lags acima de $k = q$	Declinante
ARMA(p, q)	Declinante depois de $(p-q)$ lags	Declinante para lags superiores a $(p-q)$

Fonte: Adaptado de FAVA (2000b)

Porém, esse comportamento é definido com base na FAC e na FACP teóricas. Na prática, é preciso trabalhar com as funções amostrais que não são tão bem comportadas quanto as funções teóricas.

3.3.2. ESTIMAÇÃO

Uma vez determinados os valores de p , d e q , passa-se para a estimação dos p parâmetros ϕ , dos q parâmetros θ e da variância σ_ε^2 do modelo:

$$w_t = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$
 (28)

onde $w_t = \Delta^d y_t = (1 - B)^d y_t$. (29)

A estimação pode ser feita por *mínimos quadrados* e por *máxima verossimilhança*.⁶

3.3.3. VERIFICAÇÃO

Consiste em verificar se o modelo identificado e estimado é adequado. Em caso positivo, pode-se adotá-lo para fazer previsão; em caso negativo, outra especificação

⁶ Para maiores detalhes, consultar (SOUZA & CAMARGO, 1996), (MAKRIDAKIS, et al., 1998) ou (FAVA, 2000).

deve ser escolhida para modelar a série, o que implica refazer as etapas de identificação e estimação.

As formas de verificação mais comumente consideradas são agrupadas em: *análise dos resíduos e avaliação da ordem do modelo*.

Na análise dos resíduos, verifica-se se os resíduos do modelo estimado \hat{E}_t comportam-se aproximadamente como um ruído branco. Em particular, seus coeficientes de autocorrelação devem ser estatisticamente iguais a zero. Se isso acontecer o modelo foi adequadamente especificado.

Para verificar se isso ocorre, podem ser feitos testes individuais e testes conjuntos para os coeficientes de autocorrelação $r_k(\hat{E})$, tal como na etapa de identificação.

Já na avaliação da ordem do modelo, o objetivo é verificar se o modelo não está superespecificado (p e/ou q maiores do que o devido) nem subespecificado (p e/ou q menores do que o devido).

A verificação da existência de parâmetros redundantes é feita com base no erro-padrão dos coeficientes $\hat{\phi}$ e $\hat{\theta}$ e na correlação entre eles.

Se o valor de um coeficiente estimado for pequeno em relação a seu erro-padrão, indicando sua não significância estatística, é provável que haja superespecificação. A alta correlação entre dois coeficientes do modelo indica que um deles pode ser redundante.

Para verificar se está havendo subespecificação, introduz-se parâmetros adicionais e analisa sua significância estatística.

Para maiores detalhes, consultar (MAKRIDAKIS et al., 1998) e (FAVA, 2000b).

3.3.4. PREVISÃO

Depois de selecionado o modelo, chega-se à última etapa da metodologia de Box & Jenkins, que consiste na realização de previsões para a série y_t em instantes de tempo posteriores a n .

A previsão pode ser pontual ou por intervalo. No primeiro caso, é necessário definir o previsor ótimo da série; no segundo, é preciso conhecer, adicionalmente, a distribuição do erro de previsão. Mais detalhes podem ser obtidos em (FAVA, 2000b).

3.4 MODELOS SAZONAIS

Os modelos ARIMA vistos anteriormente exploram a correlação de valores de y_t observados em instantes de tempo consecutivos.

Quando a periodicidade da série é inferior a um ano (séries mensais e trimestrais, por exemplo), outro tipo de correlação serial passa a ter importância: a correlação entre os instantes de tempo distantes entre si por s ou múltiplos de s , onde s é o número de observações contidas em um ano ($s = 12$ para dados mensais e $s = 4$ para dados trimestrais).⁷

Para acomodar esse tipo de série, surgem os modelos ARIMA sazonais, também conhecidos por modelos SARIMA, os quais são, na verdade, extensões dos modelos ARIMA.

3.4.1. MODELO SAZONAL AUTO-REGRESSIVO DE ORDEM P - SAR(P)_s

Essa classe de modelos permite apenas correlações entre instantes de tempo múltiplos de s .

$$y_t = \Phi_1 y_{t-s} + \Phi_2 y_{t-2s} + \dots + \Phi_p y_{t-ps} + \varepsilon_t \quad (30)$$

$$(1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_p B^{ps}) y_t = \varepsilon_t \quad (31)$$

$$\Phi(B^s) y_t = \varepsilon_t \quad (32)$$

A FAC do modelo SAR(P)_s, tem valores não nulos apenas para os coeficientes de autocorrelação de ordens iguais a múltiplos de s .

⁷ Embora se esteja admitindo aqui que o movimento sazonal ocorra dentro de um ano, outros períodos sazonais são também possíveis.

3.4.2. MODELO SAZONAL DE MÉDIAS MÓVEIS DE ORDEM Q - SMA(Q)_s

$$y_t = \varepsilon_t - \Theta_1 \varepsilon_{t-s} - \Theta_2 \varepsilon_{t-2s} - \dots - \Theta_Q \varepsilon_{t-Qs} \quad (33)$$

$$y_t = (1 - \Theta_1 B^s - \dots - \Theta_Q B^{Qs}) \varepsilon_t \quad (34)$$

$$y_t = \Theta(B^s) \varepsilon_t \quad (35)$$

A FAC do modelo SMA(Q)_s só apresenta valores não nulos nas ordens iguais a $s, 2s, \dots, Qs$, ou seja, ela é truncada em Qs .

3.4.3. MODELO SAZONAL AUTO-REGRESSIVO DE MÉDIAS MÓVEIS - SARMA(P, Q)_s

$$y_t = \Phi_1 y_{t-s} + \Phi_2 y_{t-2s} + \dots + \Phi_p y_{t-ps} + \varepsilon_t - \Theta_1 \varepsilon_{t-s} - \Theta_2 \varepsilon_{t-2s} - \dots - \Theta_Q \varepsilon_{t-Qs} \quad (36)$$

$$\Phi(B^s) y_t = \Theta(B^s) \varepsilon_t \quad (37)$$

A FAC deste modelo é declinante e os valores não nulos aparecem apenas quando a ordem do coeficiente de autocorrelação é um múltiplo de s .

3.4.4. MODELO SAZONAL AUTO-REGRESSIVO INTEGRADO DE MÉDIAS MÓVEIS - SARIMA (P, D, Q)_s

Esse modelo aplica-se a séries não estacionárias mas que, após a aplicação de D diferenças "sazonais", são descritas por um processo sazonal estacionário ARMA (P, Q)_s:

$$\Phi(B^s)(1 - B^s)^D y_t = \Theta(B^s) \varepsilon_t \quad (38)$$

3.4.5. MODELO SAZONAL MULTIPLICATIVO GERAL

Os modelos sazonais vistos até aqui, consideram as observações consecutivas não correlacionadas; a correlação só existe entre t e $t-s, t-2s, \dots$. A incorporação, a esses modelos, da correlação entre instantes de tempo sucessivos redundava no modelo sazonal multiplicativo geral denominado ARIMA (p, d, q) \times (P, D, Q)_s, que é dado por:

$$\begin{aligned}
 & (1 - \phi_1 B - \dots - \phi_p B^p) (1 - \Phi_1 B^s - \dots - \Phi_p B^{Ps}) (1 - B)^d (1 - B^s)^D y_t = \\
 & = (1 - \theta_1 B - \dots - \theta_q B^q) (1 - \Theta_1 B^s - \dots - \Theta_Q B^{Qs}) \varepsilon_t
 \end{aligned} \tag{39}$$

$$\phi(B) \Phi(B^s) \Delta^d \Delta_s^D y_t = \theta(B) \Theta(B^s) \varepsilon_t \tag{40}$$

3.4.6. IDENTIFICAÇÃO

O primeiro passo consiste em determinar os valores de d e de D . A análise da FAC da série original e das séries resultantes da aplicação de diferenças consecutivas e/ou sazonais indicará quais devem ser esses valores.

A determinação de p , P , q e Q é feita com base na FAC e na FACP da séries estacionarizada após a aplicação das diferenças consecutivas e/ou sazonais.

Se a série for gerada por um processo puramente sazonal ($p = q = 0$), a FAC e FACP se comportam de forma análoga às dos modelos não sazonais, com a ressalva de que os valores não nulos só ocorrem nas ordens que são múltiplos de s .

Para os modelos que contém filtros sazonais e não sazonais, a FAC e a FACP são mais complicadas.

Ilustrações gráficas das FAC's e FACP's para várias especificações de modelos podem ser encontradas em (SOUZA & CAMARGO, 1996) e (VANDAELE, 1983 in FAVA, 2000b).

3.5 CONSIDERAÇÕES FINAIS

De uma maneira geral, pode-se representar a metodologia de Box & Jenkins pela figura abaixo:

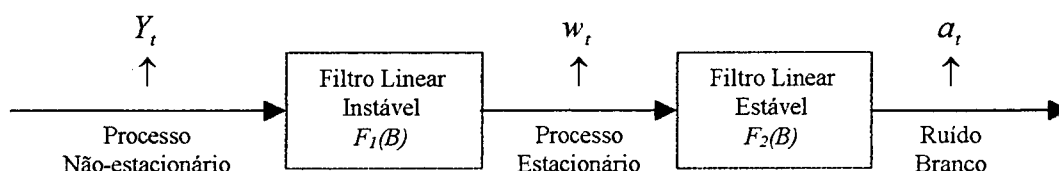


Figura 3.1 - Diagrama operacional do modelo Box & Jenkins (SOUZA & CAMARGO, 1996).

Através de uma realização Y_t de tamanho " n " do processo gerador da série, tenta-se chegar ao ruído branco a_t pela passagem sucessiva de Y_t pelos dois filtros $F_1(B)$ e $F_2(B)$ nesta ordem.

Onde:

$$F_1(B) = \nabla^d = (1 - B)^d \quad e \quad (41)$$

$$F_2(B) = \phi^{-1}(B)\theta(B) \quad (42)$$

Sendo B o operador de retardo (tal que, $B^k Y_t = Y_{t-k}$), $\theta(B)$ o polinômio das médias móveis (MA), $\phi(B)$ o polinômio auto-regressivo (AR) e ∇ o operador de diferença simples (tal que, $\nabla Y_t = Y_t - Y_{t-1}$).

É na procura de um certo conjunto de operadores diferença (passagem de Y_t para w_t) que utiliza-se a função de autocorrelação, isto é, através de *estimadores* desta função procura-se o conjunto de operadores diferença que produzam um processo cujas autocorrelações se assemelhem às correspondentes funções de algum dos processos estacionários descritos anteriormente.

O método de identificação proposto por Box & Jenkins, apesar de parecer fácil na teoria, torna-se difícil quando aplicado na prática. Conforme SOUZA & CAMARGO (1996; p. 98), *"esta dificuldade resulta do fato que se trabalha com os estimadores (ACF e PACF), que, como é de se esperar, nunca seguem exatamente o comportamento das correspondentes funções teóricas"*.

Em vez de se ter os valores teóricos das funções, dispõe-se apenas de uma realização do processo, e através desta, tenta-se obter os estimadores de tais funções.

Enfim, apesar de bem fundamentada e dos vários pacotes computacionais dedicados a previsão com a metodologia Box & Jenkins, o processo de identificação do modelo ainda é difícil (cada problema é um problema) e feito de forma manual, necessitando de muito conhecimento e prática por parte do usuário.

CAPÍTULO 4 - REDES NEURAIIS ARTIFICIAIS PARA PROCESSAMENTO TEMPORAL

*Identificar os modelos possíveis de RNAs para processamento temporal;
apontar as principais aplicações de RNAs à previsão de séries temporais.*

Esse item limita-se ao tratamento de RNAs para problemas de processamento temporal. A motivação biológica, assim como, a descrição dos principais componentes de uma RNA, dos paradigmas conexionistas (modelos) e algoritmos de aprendizado pode ser encontrada em: RUMELHART & McCLELLAND (1986); FREEMAN & SKAPURA (1992); KROSE & VAN DER SMAGT (1995); KOVÁCS (1996); LOESCH & SARI (1996); TAFNER et al. (1996); BARRETO (1997); MEHROTRA et al. (1999) e BRAGA et al. (2000).

4.1 INTRODUÇÃO

As RNA's são tipicamente usadas em reconhecimento de padrões, onde uma coleção de características (como uma imagem) é apresentada à rede e a tarefa é categorizar o padrão característico da entrada como pertencente a uma ou mais classes (RUMELHART et al., 1994), (WEIGEND & GERSHENFLD, 1994), (BRAGA et al., 2000). Outro uso típico de RNA's é em regressão não-linear, onde a tarefa é encontrar uma interpolação suave entre vários pontos (WEIGEND & GERSHENFLD, 1994). Em ambos os casos, toda a informação pertinente é apresentada à rede simultaneamente. Em contraste, a previsão de séries temporais envolve o processamento de padrões que se desenvolvem com o passar do tempo - a resposta apropriada em um ponto particular do tempo depende não apenas do valor corrente da observação, mas também dos valores prévios desta.

As RNA's tem sido um método promissor para a previsão de séries temporais desde seu surgimento (WEIGEND & GERSHENFLD, 1994). Em 1964, HU, apud WEIGEND & GERSHENFLD (1994), aplicou uma rede linear adaptável de Widrow

para o problema de previsão de tempo. Na era pós-backpropagation¹, Lapedes e Farber (1987), apud WEIGEND & GERSHENFLD (1994), treinaram uma rede - com neurônios não lineares² - para emular a relação entre a saída (o próximo valor da série) e a entrada (os valores anteriores) para séries temporais geradas por computador. Em 1990 e 1992, Weigend, Huberman e Rumelhart, apud WEIGEND & GERSHENFLD (1994, p. 25), abordaram o problema de encontrar redes de complexidade apropriada para prever séries temporais reais. Complexidade, segundo BARRETO (1997), é a quantidade de recursos em termos de RNA (topologia da rede - tamanho, estrutura e conexões) necessários para poder resolver um determinado problema, considerando-se uma certa precisão.

Segundo MOZER (1994) e BRAGA et al.(2000), uma rede que pretende prever o futuro tem que ter conhecimento sobre o passado.

De acordo com BRAGA et al. (2000), existem basicamente duas maneiras de prover memória a uma RNA:

- Introduzir atraso no tempo, como as técnicas TDNN (Time Delay Neural Network) e FIR multilayer perceptron, entre outras.
- Utilizar redes recorrentes, tais como BPTT (Back-propagation Through Time), RTRL (Real-Time Recurrent Learning), Cascade Correlation Recorrente, redes de Elman e redes de Jordan.

Visando solucionar o problema de processamento temporal, várias arquiteturas foram sendo propostas. KREMER (2000) cita vários exemplos, entre eles: o modelo de Narendra & Parthasarathy (1990); a rede TDNN (1990); rede Gama (1992); redes CasCor (1991); Memória recursiva auto-associativa (1989); Rede neural recorrente auto-associativa (1992); *Connectionist Pushdown Automaton* (1990); Máquina Conexionista de Turing (1989), e Aprendizado construtivo de Segunda-ordem (1995).

¹ Erra Pós-Backpropagation: Período após 1974, quando Paul Werbos lançou as bases do algoritmo de retro-propagação (backpropagation).

² Segundo Barreto (1997, p.67), a não linearidade do neurônio frequentemente é introduzida na função logística, tangente hiperbólica ou em degraus.

Com o intuito de classificar os diversos modelos de redes neurais para processamento temporal, HORNE & GILES (1995) apud KREMER (2000), em um artigo que compara o desempenho de diferentes arquiteturas para previsão de séries de tempo, desenvolveram uma taxinomia simples para processamento temporal. O método divide o universo de arquiteturas existentes em dois. A primeira partição separa as redes cujas representações de estado é codificada em unidades de entrada e saída (redes com estados observáveis) das redes cujas representações de estado é codificada nas unidades ocultas (redes com dinâmica oculta).

Redes com "estados observáveis" incluem a rede de Narendra e Parthasarathy (1990), a rede TDNN de Lang, Waibel e Hinton (1990), e as redes Gama de Vries & Príncipes (1992). A classe de redes com "dinâmica oculta" é dividida em redes de uma única camada, multi-camadas e redes com realimentação local. A taxinomia de Horne & Giles é ilustrada no Quadro 4.1.

Quadro 4.1 - Taxinomia de Horne & Giles (KREMER, 2000)

Redes com Estados Observáveis
Narendra & Parthasarathy
TDNN (Lang, Waibel & Hinton)
Rede Gama (De Vries & Príncipe)
Dinâmica Oculta
<i>Uma-Camada</i>
Primeira Ordem
Ordem Superior (Giles et al.)
Bilinear
Quadrática (Watrous & Kuhn)
<i>Multi-Camadas</i>
Robinson & Fallside
Redes Recorrentes Simples (Elman)
<i>Realimentação Local</i>
Back & Tsoi
Fransconi, Gori & Soda
Poddar & Unnikrishnan

TSOI & BACK (1994) apud KREMER (2000) desenvolveram uma taxinomia especificamente projetada para redes localmente recorrentes e globalmente diretas, um subconjunto de redes neurais para processamento temporal. Nessas redes, todas as conexões são diretas com exceção de uma conexão local temporal para cada nodo. A taxinomia de Tsoi & Back baseia-se no tipo de retardo incorporado na conexão

recorrente (tipo de sinapse) e no valor de retardo (localização do ciclo). A "localização do ciclo" é subdividida em três dimensões, dependendo da combinação dos valores dos nós anteriores; da entrada dos nós anteriores; ou do valor prévio transmitido pelas sinapses. A taxinomia de Tsoi & Back é representada no Quadro 4.2.

Quadro 4.2 - Taxinomia de Tsoi & Back (KREMER, 2000)	
Tipo de Sinapse	
-	Simple
-	Dinâmica
Realimentação nas Sinapses	
-	Sim
-	Não
Realimentação na Ativação	
-	Sim
-	Não
Realimentação na Saída	
-	Sim
-	Não

MOZER (1994) desenvolveu uma taxinomia para processamento temporal que baseia-se na suposição de que toda rede neural para processamento temporal consiste em dois elementos: uma memória de curto-tempo e um elemento de previsão. A memória de curto-tempo calcula o estado da rede, enquanto o elemento de previsão usa o estado para calcular a saída da rede. ZHANG & HUTCHINSON (1994) e DE VRIES & PRINCIPE (1992) apud MOZER (1994), também fazem a distinção entre estes dois componentes. Mozer classificou a memória de curto-tempo ao longo de três dimensões: *conteúdo* e *forma* (que define de que forma a memória de curto-tempo é calculada) e *adaptabilidade* (que define como as mudanças na memória são feitas durante o processo de adaptação). Segundo Mozer, o elemento de previsão para processamento temporal é sempre uma rede neural direta. O Quadro 4.3 ilustra a taxinomia de Mozer.

Quadro 4.3 - Taxinomia de Mozer.

Conteúdo
- Input (I)
- Transformed Input (TI)
- Transformed Input and State (TIS)
- Output (O)
- Transformed Output (TO)
- Transformed Output and State (TOS)
Forma
- Delay-Line
- Exponencial
- Gamma
Adaptabilidade
- Estática
- Dinâmica

Apesar de só se aplicar a memória de curto-tempo da rede (Mozer não descreve o elemento de previsão, apenas diz que esta é uma rede neural direta), a taxinomia de Mozer é ainda a mais sofisticada, compreensível e utilizada para o processamento temporal (KREMER, 2000).

4.2 TAXINOMIA DE MOZER

Segundo MOZER (1994), ao projetar uma RNA para o problema de previsão de séries temporais, alguns aspectos devem ser observados:

ARQUITETURA

Qual deve ser a estrutura interna da memória de curto-tempo e do componente preditor? Responder a esta pergunta, segundo MOZER (1994), envolve especificar o número de camadas e de unidades da rede, o padrão de conectividade entre as unidades, e a dinâmica de ativação dessas.

TREINAMENTO

Dado um conjunto de exemplos de treinamento, como os parâmetros internos do modelo (os pesos das conexões) devem ser ajustados? Cada exemplo de treinamento i consiste de uma série de entradas, $\{x_i(1), x_i(2), x_i(3), \dots, x_i(t_i)\}$, e uma série de previsões associada, $\{p_i(1), p_i(2), p_i(3), \dots, p_i(t_i)\}$, onde t_i é o número de instantes do exemplo e $p_i(\tau)$ é o valor previsto desejado no momento

τ . Treinar um modelo de RNA envolve fixar seus pesos tal que suas previsões, $y_i(\tau)$, aproximem-se o máximo possível das previsões alvo, $p_i(\tau)$, normalmente no sentido dos mínimos quadrados.

REPRESENTAÇÃO

Na representação procura-se responder a pergunta: Como a série temporal será representada na memória de curto-tempo? A natureza e a quantidade da informação de entrada dependem do domínio.

A arquitetura, a dinâmica da rede, o procedimento de treinamento e a representação dos dados estão completamente relacionados e, por isso, podem ser vistos sob perspectivas diferentes em problemas semelhantes (MOZER, 1994). Uma determinada escolha de representação exige uma certa arquitetura ou um tipo particular de algoritmo de aprendizado para computar a representação. De modo oposto, a arquitetura ou o algoritmo de aprendizado podem restringir a classe de representações possíveis.

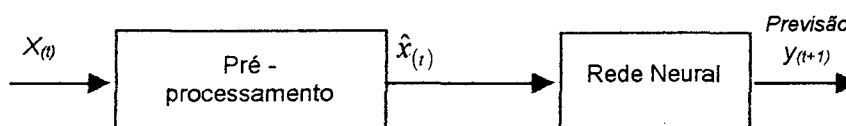


Figura 4.1 - Formulação abstrata da tarefa de previsão temporal. (Adaptado de MOZER (1994)).

A Fig. 4.1, apresenta o modelo básico de uma RNA para o problema de previsão, segundo MOZER (1994). A entrada no momento t é denotada por $x_{(t)}$. Para séries univariadas, essa entrada é um valor escalar; para séries multivariadas, x deve ser um vetor valorado. A saída do componente preditor no momento t , $y_{(t)}$ é baseada na seqüência de entrada $\hat{x}_{(1)} \dots \hat{x}_{(t)}$ e representa um valor futuro da série de entrada ou, possivelmente, alguma função do valor futuro; y também pode ser um vetor multivalorado. A memória de curto-tempo (ou pré-processamento feito nos dados) guarda os aspectos da seqüência de entrada relevantes à tarefa de previsão.

MOZER (1994) abordou o assunto da representação caracterizando o universo de modelos de memórias de curto-tempo de RNAs. Segundo o autor, há três dimensões ao longo das quais os modelos de memória variam: quanto a forma de memória, o conteúdo, e a adaptabilidade.

4.2.1. FORMAS DE MEMÓRIA DE CURTO-TEMPO

4.2.1.1. MEMÓRIA TAPPED DELAY-LINE

A forma mais simples de memória é uma memória contendo as n entradas mais recentes. Tal memória é freqüentemente chamada de *Tapped Delay-line*, porque pode ser formada por uma série de linhas de atraso. É também chamada de *Delay Space Embedding*, e forma a base dos modelos estatísticos autoregressivos (AR) tradicionais. Esse tipo de memória é muito comum em modelos de RNA's (ex. BRAGA et al., 1996, 1999; YANG, et al., 1996; STERN, 1996; FARAWAU & CHATFIELD, 1998).

Uma memória do tipo *Tapped Delay-line*, equivale a selecionar certos elementos de uma sequência $x(1) \dots x(t)$, ou seja, um total de Ω elementos, e formar um estado de representação $(\bar{x}_1(t), \bar{x}_2(t), \bar{x}_3(t), \dots, \bar{x}_\Omega(t))$, onde $\bar{x}_i(t) = x(t - i + 1)$ (Fig. 4.3).

Para esse tipo de memória, pré-processar consiste meramente em armazenar valores passados da variável e os prover à rede, junto com o valor mais recente.

Uma extensão secundária desta formulação, para permitir amostras de valores passados não uniformes, envolve especificar atrasos variáveis, como $\bar{x}_i(t) = x(t - w_i)$, onde w_i é o retardo (inteiro) associado ao componente i .

Vários tipos de pré-processamento para problemas de predição podem ser descritos como uma transformação da sequência de entrada por uma "função núcleo". Se existe um valor c_i para cada $i = 1, 2, \dots, n$, o $i^{\text{ésimo}}$ componente do vetor transformado é determinado pela equação:

$$\bar{x}_i(t) = \sum_{\tau=1}^t c_i(t - \tau)x(\tau), \quad (43)$$

Uma memória do tipo *Delay-line* é um caso especial da Equação 43, em que a função núcleo é descrita por:

$$c_i(t) = \begin{cases} 1, & \text{se } t = w_i; \\ 0, & \text{se } t \neq w_i. \end{cases} \quad (44)$$

A Fig. 4.2 mostra a função núcleo. Substituindo esta por uma função núcleo diferente, obtém-se as outras formas de memória de curto-tempo que serão apresentadas a seguir.

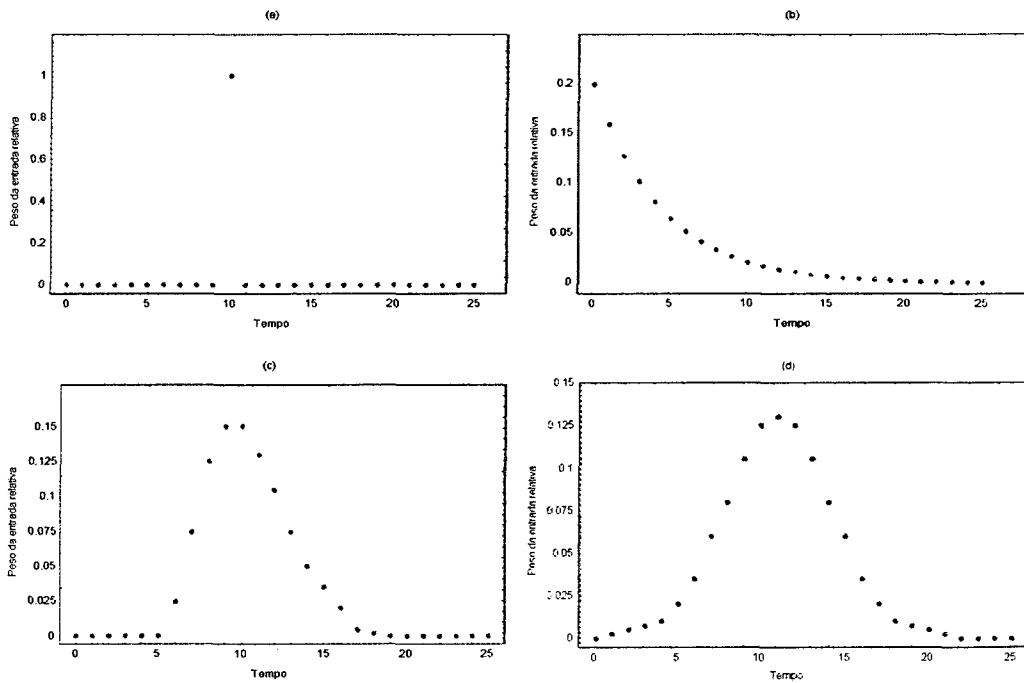


Figura 4.2 - Função núcleo para (a) uma memória *Delay-line*, $w = 10$; (b) uma memória *Exponencial*, $\mu = 0,8$; (c) uma memória *Gama*, $w = 6$ e $\mu = 0,4$, e (d) uma memória *Gaussiana* (MOZER, 1994).

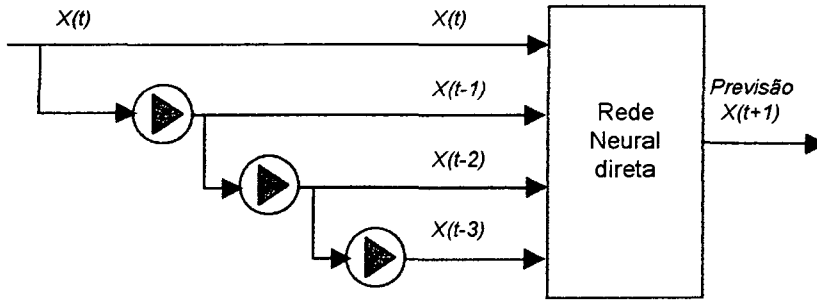


Figura 4.3 - Modelo de memória Tapped delay-line, segundo a Taxinomia de Mozer. (Adaptado de MOZER (1994)).

4.2.1.2. MEMÓRIA EXPONENCIAL

Uma memória do tipo exponencial é formada usando-se a função núcleo:

$$c_i(j) = (1 - \mu_i) \mu_i^j \quad (45)$$

onde μ_i pertence ao intervalo $[-1, 1]$ (Fig. 4.2 (b)). Ao contrário da memória *Delay-line*, a força de uma entrada não decai bruscamente em um ponto fixo do tempo, e sim, exponencialmente. Assim sendo, valores de entrada mais recentes terão sempre uma força maior do que os valores mais distantes. Todavia, na ausência de ruído, uma memória exponencial pode preservar toda a informação disponível em uma seqüência (MOZER, 1994).

As variáveis de entrada transformadas podem ser atualizadas incrementalmente pela equação:

$$\bar{x}_i(t) = (1 - \mu_i) x_i(t) + \mu_i \bar{x}_i(t-1) \quad (46)$$

Com $\bar{x}_i(0)$ igual a 0. Essa regra é um caso especial do modelo de Médias Móveis conhecido como MA(1) na literatura estatística.

4.2.1.3. MEMÓRIA GAMA

A memória Gama é uma generalização da memória *Delay-line* e da memória Exponencial, que permite ao usuário variar a *profundidade* (quantidade de informação passada que a memória pode armazenar) como também a *resolução* (grau de informação,

relativa aos valores individuais da seqüência de entrada, que é preservada). A função núcleo de uma memória Gama para um intervalo de tempo discreto é:

$$c_i(j) = \begin{cases} (j/l_i)(1-\mu_i)^{l_i+1} \mu_i^{j-l_i} & \text{se } j \geq l_i; \\ 0 & \text{se } j < l_i. \end{cases} \quad (47)$$

onde o retardo l_i é um inteiro não negativo e $\mu_i \in [0,1]$ (Fig. 4.2 (c)). Este modelo reduz-se a uma memória exponencial quando $l_i = 0$, e a uma memória *Delay-line* quando μ_i aproxima-se de 0. As variáveis de entrada transformadas são atualizadas na memória Gama pela equação:

$$\bar{x}_{\mu,j}(t) = (1-\mu)\bar{x}_{\mu,j-1}(t-1) + \mu\bar{x}_{\mu,j}(t-1) \quad (48)$$

com

$$\bar{x}_{\mu,-1}(t) = x(t+1) \quad \text{para } t \geq 0, \quad e \quad \bar{x}_{\mu,j}(0) = 0 \quad \text{para } j \geq 0. \quad (49)$$

4.2.1.4. OUTRAS FORMAS DE MEMÓRIAS

De acordo com MOZER (1994), qualquer função núcleo utilizada resulta em uma forma de memória distinta. Por exemplo, uma função núcleo Gaussiana (Fig. 4.2 (d)) pode ser usada para obter uma memória simétrica em volta de um determinado ponto no tempo. O que torna a memória Gama e seus casos especiais particularmente úteis é que estas podem ser computadas por um procedimento de atualização incremental, enquanto que formas alternativas, como a memória Gaussiana, requerem a avaliação da transformação do núcleo com a seqüência de entrada a cada passo de tempo.

4.2.2. CONTEÚDO DA MEMÓRIA DE CURTO-TEMPO

Embora uma memória tenha que guardar informações relativas a seqüência de entrada, esta não precisa, necessariamente, armazenar toda a seqüência na íntegra (MOZER, 1994). O processo de codificação da memória pode incluir um passo adicional no qual a sucessão de entrada $x(1)\dots x(t)$ é transformada em uma nova representação

$x'(1) \dots x'(t)$, e é esta representação transformada que é codificada na memória. Assim, as entradas \bar{x}_i são definidas em termos de x' , e não de x .

Dessa forma, quando ao conteúdo, uma memória pode ser classificada em:

MEMÓRIA-I (INPUT MEMORY)

$$x'(\tau) = x(\tau). \quad (50)$$

ou seja, o valor de x é transformado pela função identidade.

MEMÓRIA-TI (TRANSFORMED INPUT MEMORY)

$$x'(\tau) = f(x(\tau)). \quad (51)$$

Neste caso, x é transformado por uma função vetorial f , não-linear. Geralmente f é a função de ativação padrão da rede, que calcula a soma ponderada dos elementos de entrada e passa esse valor por uma função não-linear, como por exemplo a função tangente hiperbólica (Equação 61, p. 58).

MEMÓRIA-TIS (TRANSFORMED INPUT AND STATE MEMORY)

A transformação não-linear pode ser feita não só em cima da sequência de entrada, mas também, sobre o estado interno atual da memória.

$$x'(\tau) = f(x(\tau), \bar{x}_1(\tau), \dots, \bar{x}_\Omega(\tau)). \quad (52)$$

Esse tipo de memória pode ser implementada em uma arquitetura de rede neural recorrente na qual \bar{x}_i e x' correspondem ao processamento em duas camadas ocultas da rede (Fig. 4.4 (a)). Esta arquitetura é similar a uma rede neural recorrente padrão, com processamento sequencial (Fig. 4.4 (b); ex., Mozer (1989); Elman (1990)).

Para tarefas de auto-previsão na qual a saída desejada, $p(\tau)$, é uma previsão um-passo-à-frente da entrada; por exemplo:

$$p(\tau) = x(\tau + 1), \quad (53)$$

pode-se considerar um conteúdo alternativo para a memória. Em lugar de alimentar a rede com o valor atual da sequência, a previsão anterior pode ser usada.

$$x'(\tau) = p(\tau - 1). \quad (54)$$

Esta transformação é chamada **MEMÓRIA-O (OUTPUT MEMORY)**. Análogo as memórias TI e TIS, memórias **TO (TRANSFORMED OUTPUT MEMORY)** e **TOS (TRANSFORMED OUTPUT AND STATE MEMORY)** também podem ser construídas, sugerindo uma caracterização do conteúdo da memória ao longo de duas subdimensões: 1) o tipo de transformação aplicada e 2) a que valores esta transformação é aplicada (valores de entrada ou de saída da rede).

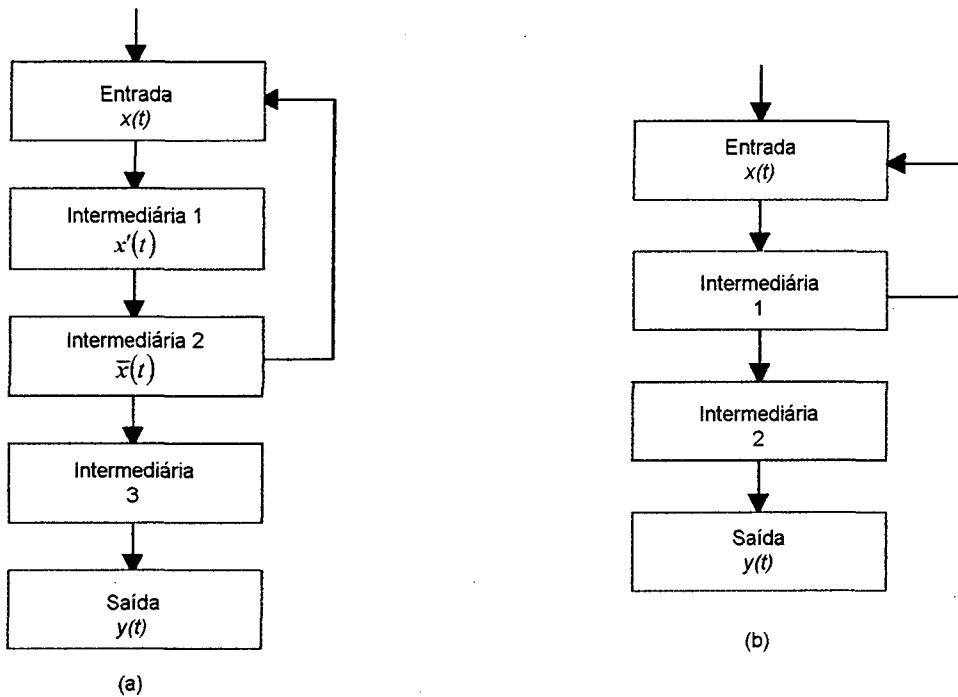


Figura 4.4 - (a) Uma memória TIS em uma arquitetura de rede neural; (b) Arquitetura padrão de uma rede neural recorrente (Adaptado de MOZER (1994)).

Na Figura 4.4, cada retângulo representa uma camada de neurônios, e cada seta representa conectividade completa de uma camada para a outra. A camada de entrada corresponde ao elemento corrente da série temporal, $x(t)$; a primeira camada intermediária corresponde a representação TIS, $x'(t)$; e a segunda camada intermediária corresponde ao estado da memória, $\{\bar{x}_i(t)\}$. Segundo MOZER (1994), uma arquitetura padrão recorrente, é semelhante a arquitetura TIS exceto que as primeiras duas camadas intermediárias estão concatenadas; ex.: $x'(t) = [\bar{x}_1(t) \dots \bar{x}_\Omega(t)]$.

As Figuras de 4.5 à 4.9, ilustram o tipo de dinâmica aplicada nos dados, e o Quadro 4.4, apresenta as transformações possíveis.

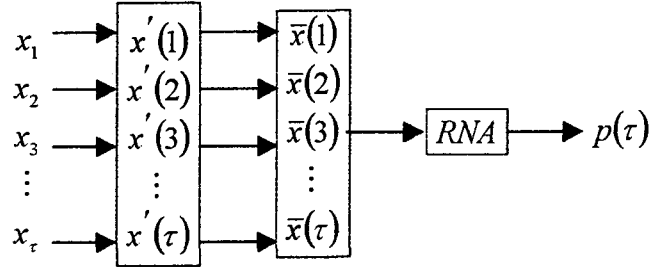


Figura 4.5 - Arquitetura da memória I e TI.

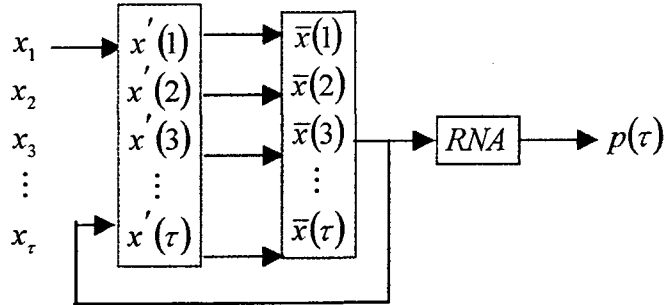


Figura 4.6 - Arquitetura da memória TIS.

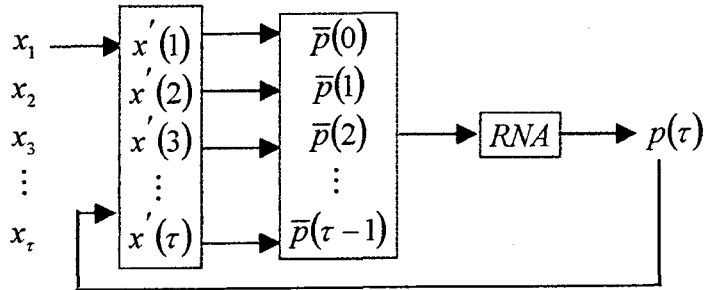


Figura 4.7 - Arquitetura da memória O e TO.

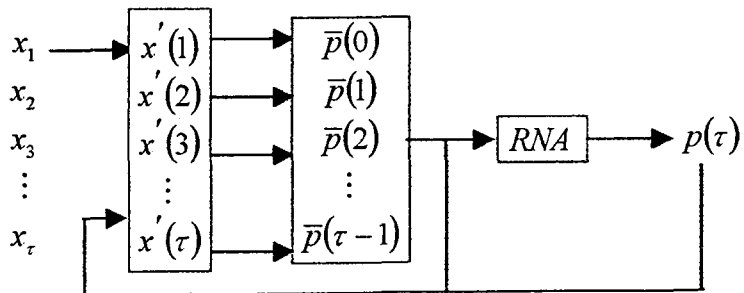


Figura 4.8 - Arquitetura da memória TOS.

Quadro 4.4 - Resumo dos principais tipos de transformações e formas de memória.

Tipo de Transformação	
I	$x'(\tau) = x(\tau)$
TI	$x'(\tau) = f(x(\tau))$
TIS	$x'(\tau) = f(x(\tau), \bar{x}_1(\tau), \dots, \bar{x}_n(\tau))$
O	$x'(\tau) = p(\tau - 1)$
TO	$x'(\tau) = f(p(\tau - 1))$
TOS	$x'(\tau) = f(p(\tau - 1), \bar{p}_1(\tau - 1), \dots, \bar{p}_n(\tau - 1))$
Forma da Memória	
$\bar{x}_i(t) = \sum_{\tau=1}^t c_i(t - \tau) x'(\tau)$	
Tapped Delay-Line	$c_i(t) = \begin{cases} 1, & \text{set} = w_i; \\ 0, & \text{set} \neq w_i. \end{cases}$
Exponencial	$c_i(j) = (1 - \mu_i) \mu_i^j,$
Gama	$c_i(j) = \begin{cases} (j/l_i)(1 - \mu_i)^{l_i+1} \mu_i^{j-l_i} & \text{se } j \geq l_i; \\ 0 & \text{se } j < l_i. \end{cases}$

4.2.3. ADAPTABILIDADE DA MEMÓRIA

MOZER (1994) ainda classifica as memórias de curto-tempo conforme a adaptabilidade. Uma memória tem vários parâmetros - $\{\mu_i\}, \{\Omega_i\}$, e o k da Equação 61, p. 58 - que devem ser especificados. Se os parâmetros são fixados com antecedência, a memória é estática. O estado da memória, $\{\bar{x}_i(t)\}$, é uma função predeterminada da sequência de entrada, $x(1) \dots x(t)$. Neste caso, a tarefa da rede neural é fazer as melhores previsões possíveis dado a representação fixa do histórico de entrada. Em contraste, se a rede neural pode ajustar os parâmetros da memória, a representação da memória é adaptável. Essencialmente, ajustando os parâmetros da memória, a rede neural seleciona, dentro dos limites da capacidade da memória, que aspectos da sequência de entrada estão disponíveis para fazer as previsões. Além de aprender a fazer previsões, a rede neural tem que aprender também as características da memória - especificadas pelos parâmetros - que melhor facilitam a tarefa de previsão (MOZER, 1994).

Casos interessantes de modelos de memórias adaptativas na literatura de RNA incluem:

- A aprendizagem dos retardos (Bodenhausen & Waibel (1991); e Unnikrishnan, Hopfield & Tank (1991)) apud MOZER (1994);
- a aprendizagem das taxas de decadência exponenciais (Bachrach (1988); Frasconi, Gori & Soda (1992); e Mozer (1989)) apud MOZER (1994);
- a aprendizagem dos parâmetros da memória gama (De Vries & Principe (1992) apud MOZER (1994));
- toda arquitetura de rede neural treinada com backpropagation através do tempo - BPTT (Rumelhart, Hinton & Williams (1986) apud BRAGA et al. (2000)) ou,
- e o aprendizado recorrente de tempo-real - RTRL (Robinson & Fallside (1987); Schmidhuber (1992); e Williams & Zipser (1989)) apud BRAGA et al. (2000), onde os k_s da Equação 61, pg. 58, são ajustados durante o treinamento.

O algoritmo SRN de Elman, segundo BRAGA et al., 2000, se encaixa em algum lugar entre uma memória adaptável e uma estática devido ao procedimento de treinamento.

Modelos estáticos de memória podem fornecer uma aproximação razoável se houver um conhecimento adequado do domínio para compelir o tipo de informação que deve ser preservada na memória (MOZER, 1994).

O Quadro 4.5 apresenta algumas combinações de conteúdo e forma de memória, juntamente com os modelos de cada classe, existentes na literatura. A memória *I-Delay* é a mais simples, e corresponde a uma rede neural direta com retardos na apresentação dos dados de entrada.

Quadro 4.5 - Arquitetura de RNA para processamento temporal (adaptada de MOZER (1994))

Formas de memória			
Conteúdo	Delay-line	Exponencial	Gama
I	ZHANG & HUTCHINSON, 1993; YANG, et al., 1996; STERN, 1996; LAPEDES & FARBER, 1987; ELMAN & ZIPSER, 1988; FARAWAY & CHATFIELD, 1998; BRAGA et al., 1996, 1999	?	DE VRIES & PRINCIPE, 1991, 1992
TI	SOMPOLINSKY & KANTER, 1986; KLEINFELD, 1986; WAIBEL et al., 1989; WAN, 1993	BACHRACH, 1988; MOZER, 1989; FRASCONI, GORI, & SODA, 1992	?
TIS	HERZ, 1991	MOZER, 1992	?
O	CONNOR, ATLAS, & MARTIN, 1992	JORDAN, 1987	?

Além das classes simples de memória, arquiteturas híbridas também tem sido propostas. MOZER (1991) apud MOZER (1994), por exemplo, propôs arquiteturas combinando uma memória *I-Delay* com uma *TIS-Exponencial*, e uma memória *I-Delay* com uma *TI-Exponencial*.

4.3 ARQUITETURA PERCEPTRON MÚLTIPLAS CAMADAS

A arquitetura *Perceptron* Múltiplas Camadas (*Multilayer Perceptron - MLP*) com aprendizado baseado no algoritmo de retropropagação é o paradigma de RNA de propósito mais geral e mais utilizado em áreas como reconhecimento de padrões, processamento de fala e previsão de séries temporais.

A base teórica deste modelo foi apresentada originalmente em 1974 por Paul Werbos, e redescoberto, independentemente, por D. Parker (1982), Le Cun (1985) e D. Rumelhart et al. (1986) (DE AZEVEDO, 1999).

O modelo do *perceptron* multi-camadas utiliza uma topologia de três ou mais camadas. As conexões entre as unidades são do tipo intercamadas e são direcionadas da camada de entrada para a camada de saída (Figura 4.9).

No modelo dos *perceptrons* multi-camadas, entre uma camada e outra, existe uma matriz de pesos. A regra de propagação é a combinação entre as saídas de cada

unidade e a matriz de pesos. Ela é realizada através da soma ponderada de cada sinal que chega, via conexões, pelo respectivo peso. O estado de ativação assume valores contínuos e, devido a isto a regra de ativação das unidades utiliza como função de ativação uma função contínua - funções semilineares (BRAGA et al, 2000).

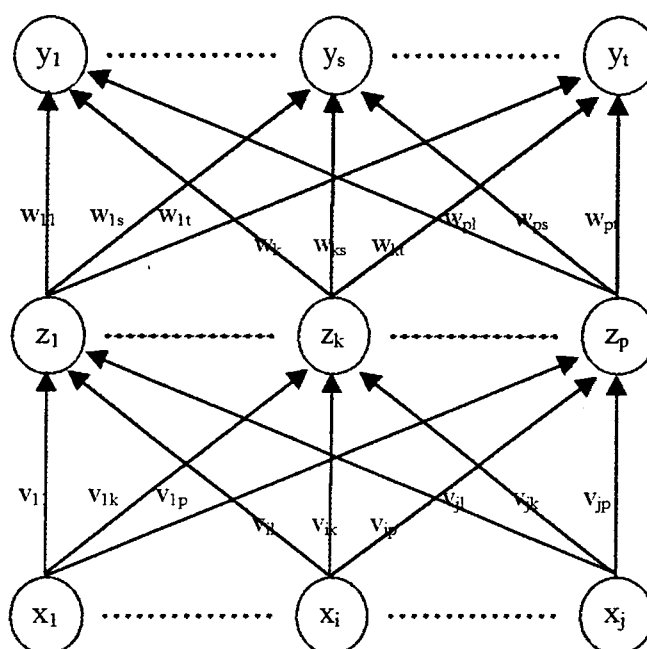


Figura 4.9 - Topologia genérica do Perceptron Multicamadas

Uma função semilinear também se faz necessária pela regra de aprendizado que utiliza uma função de ativação contínua, não-decrescente e diferenciável.

A rede MLP pode ser vista como uma aplicação do método de otimização conhecido como aproximação estocástica (DE OLIVEIRA, 1999). Ela realiza aproximações globais para o mapeamento entre a entrada e a saída da rede. O treinamento da rede é normalmente efetuado por meio do algoritmo de Retropropagação (*Backpropagation*). Este algoritmo executa um treinamento supervisionado da rede, utilizando a retropropagação do erro entre a saída produzida pela rede e o alvo. O treinamento é feito em três fases: propagação à frente (*feedforward*) do padrão de entrada, cálculo e retropropagação do erro apurado e ajuste dos pesos associados às conexões dos neurônios na direção do menor erro entre a saída produzida pela rede e o alvo. Após o treinamento, o uso da rede envolve apenas a fase de propagação à frente.

A base matemática para o algoritmo de retropropagação é a técnica de otimização do gradiente descendente. A idéia básica é calcular derivadas parciais de uma função de aproximação $f(w;x)$, realizada pela rede para todos os elementos do vetor de pesos ajustáveis W para um dado valor do vetor de entrada x (DE OLIVEIRA, 1999). O gradiente da função dá a direção na qual a função diminui mais rapidamente. Aplicada a minimização de erro da rede, a função representa o erro e suas variáveis são os pesos da rede.

Os parâmetros livres, que necessitam ser ajustados para o treinamento da rede são: o número de neurônios na camada intermediária, a inicialização dos pesos da rede, o valor da taxa de aprendizado, o erro máximo admissível no treinamento e o número máximo de épocas de treinamento. Caso não haja conhecimento prévio a respeito de quais sejam os bons valores de inicialização de pesos para o problema em questão, uma boa alternativa é a inicialização aleatória. A taxa de aprendizado se traduz ao tamanho dos passos dados, ao percorrer a superfície de erro, em busca do ponto de erro mínimo global. Passos mais largos propiciam um treinamento mais rápido, porém aumentam o risco de que se passe o ponto de mínimo e a rede não consiga estabilizar. O número de épocas de treinamento e o erro máximo admissível no treinamento são critérios para a parada do treinamento. Redes treinadas com um número de épocas excessivo memorizam os vetores de entrada, perdendo a capacidade de generalização. Por fim, um valor muito baixo para o erro máximo admissível pode se mostrar impossível de ser alcançado durante o treinamento.

Um outro parâmetro que pode ser utilizado no algoritmo de retropropagação é o termo de momento (*momentum*). A introdução deste parâmetro modifica a direção para a alteração de pesos. A direção que era em função do gradiente descendente corrente, passa a ser em função da combinação entre os gradientes descendentes corrente e anterior. Esta técnica procura aumentar as chances da rede estabilizar no ponto de erro mínimo global, diminuindo portanto a possibilidade da rede estabilizar em um ponto de mínimo local.

A rede MLP treinada com o algoritmo de retropropagação possui equilíbrio entre uma boa memorização do conjunto de treinamento e uma boa capacidade de resposta às entradas similares ao padrão de treinamento (generalização). Isto habilita a rede a reconhecer padrões que são similares aos do conjunto de treinamento ou padrões com ruído. Suas desvantagens são o elevado tempo para treinamento e o grande número de parâmetros a ajustar. A quantidade de parâmetros soma-se à falta de uma metodologia para o ajuste ideal de seus valores. Os valores são geralmente obtidos através de técnicas heurísticas e experimentação.

Uma descrição detalhada da arquitetura Perceptrons Múltiplas Camadas pode ser encontrada em (FREEMAN & SKAPURA, 1992), (TAFNER et al., 1996), (LOESCH & SARI, 1996), (KOVÁCS, 1996) e (BRAGA et al., 2000).

4.4 ARQUITETURA DE REDES NEURAIIS RECORRENTES PARA PREVISÃO DE SÉRIES DE TEMPO

Segundo MOZER (1994), a arquitetura "padrão" de RNA recorrentes para processamento seqüencial equívale a uma memória de curto-tempo TIS-Exponencial com $\mu_i = 0$ ou, uma memória TIS-Delay com $\Omega = 0$.

Uma arquitetura TIS-0 parece, a princípio, poderosa o suficiente para lidar com tarefas arbitrárias. Com um número grande de unidades ocultas e bem treinada, a arquitetura deveria ser capaz de reter a informação necessária para a tarefa pertinente. Isso é alcançado pelo ajuste dos parâmetros w , das Equações 52 e 53, através do procedimento de descida do gradiente. Na prática, porém, estudos têm concluído que o procedimento de descida do gradiente não é suficientemente poderoso para descobrir os tipos de relações que existem em uma seqüência temporal, especialmente as relações que duram um período de tempo mais longo e aquelas envolvendo estatísticas de ordem extremamente alta (WEIGEND & GERSHENFLD, 1994).

BENGIO et al. (1993) apud MOZER (1994) apresentam argumentos teóricos para as limitações inerentes da etapa de aprendizagem em RNA recorrentes. MOZER (1989, 1992) apud MOZER (1994) ilustra empiricamente que tarefas relativamente

simples, não podem ser resolvidas por memórias do tipo TIS-0. Embora memórias TIS-0 tenham se mostrado inadequadas para estas tarefas, memórias TI-Exponencial ou TIS-Exponencial tem apresentado um desempenho significativo. Em ambos os casos, o componente exponencial permitiu às redes atravessar em intervalos maiores de tempo.

Com arquiteturas do tipo TIS-0, o método de descida do gradiente frequentemente não é capaz de encontrar valores significativos para os pesos das conexões recorrentes (MOZER, 1994).

4.5 APLICAÇÕES DE RNAs À PREVISÃO

Devido à sua capacidade de manipular dados imprecisos e ruidosos, as RNAs vêm se tornando uma poderosa ferramenta para processamento de informações econômicas, muitas vezes com vantagens sobre vários métodos estatísticos, essencialmente lineares. Muitas aplicações têm sido desenvolvidas nos últimos anos na área econômica, sendo que a maioria destas aplicações têm mostrado bons resultados quando comparados com procedimentos estatísticos. Um bom exemplo desta superioridade está no estudo feito por CHAKRABORTY et al. (1992), onde os autores utilizaram RNAs para fazer previsões univariadas e multivariadas no preço de grãos em três cidades americanas: Buffalo, Minneapolis e Kansas. A rede utilizada foi uma rede direta *I - Delay* utilizando o algoritmo Backpropagation. Os autores utilizaram dados do período de 1972 a 1980 e os resultados das RNAs foram comparados com os obtidos pelo modelo de Tiao & Tsay, comprovando a eficiência das RNAs.

Outro exemplo bem sucedido de aplicação das RNAs na previsão de séries econômicas foi feito por PODDIG (1993), onde se utilizou RNAs para prever as mudanças na taxa de câmbio USD/DM, um, três e cinco dias à frente. O autor baseou-se em estudos da *SIEMENS Corporate Research* e utilizou 44 séries de indicadores fundamentais e técnicos como produto de consumo, metais preciosos, indicadores do mercado financeiro e moedas estrangeiras, entre outros, para treinar as RNAs.

Poddig dividiu suas análises em duas fases. Na primeira fase ele utilizou redes com uma camada e redes MLP com duas camadas (uma intermediária), que foram

treinadas com o algoritmo de retropropagação. O autor fez comparações entre os resultados das duas redes e entre elas e os métodos estatísticos baseados em análise de fatores discriminantes, tendo as RNAs mostrado melhores resultados.

Na segunda fase da análise, o autor usou redes RCP (*ReCurrent Perceptron*) em busca de resultados com menor flutuação e maior estabilidade. Desta feita, Poddig comparou os resultados das RNAs com métodos de regressão multivariada, obtendo novamente melhores resultados para as RNAs.

Outra aplicação interessante de RNAs na previsão de taxa de câmbio foi feita por REFENES (1992), onde os autores fizeram previsões "single" e "multi-step" utilizando o algoritmo de retropropagação para prever a taxa de câmbio US/DM. Os autores analisaram os principais fatores no projeto das RNAs, como arquitetura, função erro, taxa de aprendizado e função de ativação. Além disso, os autores discutem o efeito de diferentes funções de ativação na convergência das RNAs, mostrando que, para o caso em questão, as funções simétricas (*tanh*) convergem mais rápido que as funções assimétricas (*sigmoide*). Por fim, os autores comparam os resultados com modelos estatísticos de alisamento exponencial e Box & Jenkins puramente autoregressivo, obtendo resultados bastante satisfatórios em uma carteira de investimentos, com no mínimo 22% de lucro contra 2-5% dos métodos convencionais.

Outras aplicações bem sucedidas foram realizadas na previsão de séries temporais, tanto na área econômica como em outras áreas. Entre elas pode-se destacar séries como: "Índice de passageiros de linhas aéreas" (FERNANDES, 1995), (PORTUGAL, 1995) e (FARAWAY & CHATFIELD, 1998), "Previsão de vazões" (KADOWAKI et al., 1997), (VALENÇA & VASCONCELOS, 1997) e (LIMA Jr. & DE CARVALHO, 1997), "Preço de ações" (MUELLER, 1996) entre outras, mostrando que o interesse pela área tem crescido bastante.

O Quadro 4.6 mostra um resumo das principais aplicações de RNAs na previsão de séries temporais, indicando a série utilizada, o tipo de rede, o algoritmo de treinamento, os métodos de avaliação de desempenho utilizados e os resultados obtidos.

Quadro 4.6 - Resumo das principais aplicações de RNAs na previsão de séries temporais.

Série de Tempo	Memória de Curto-Tempo	Componente Preditor	Medidas de Erro	Métodos de Comparação	Resultados Obtidos
<p>A - Índice mensal de passageiros de linhas aéreas nos EUA.</p> <p>B - Índice pluviométrico anual da cidade de Fortaleza.</p> <p>C - Produto industrial do estado do RS. (FERNANDES, 1995)</p> <p>Série C (PORTUGAL, 1995)</p>	Metodologia de Varfis e Versino (1990)	Rede Neural Direta com <i>Backpropagation</i>	MAE RMSE	<p>Modelo Estrutural Clássico - UCM (FERNANDES, 1995)</p> <p>BOX & JENKINS e Modelo Estrutural Clássico - UCM (PORTUGAL, 1995)</p>	<p>RNA superiores para horizontes longos de previsão</p> <p>Modelo Estrutural superior para Previsões um-passo-a-frente.</p> <p>Série A RMSE = 0,6271</p> <p>Série B 19,12 para 12 passos à frente e Série C 4,17 para 7 passos à frente.</p> <p>RMSE (7 passos à frente)</p> <p>BOX & JENKINS = 9,96</p> <p>UCM = 5,49</p> <p>RNA = 5,73</p>
Preço do café pago ao produtor. (BRAGA et al., 1999)	I - Delay	Rede Neural Direta com Rprop (<i>Resilient Backpropagation</i>)	PE	BOX & JENKINS	Média do erro para RNA ficou em 6,1% enquanto que para o Modelo BOX & JENKINS ficou em 4,7%.
Sistema de drenagem subterrâneo. (YANG et al., 1996)	Simples e I - Delay	Rede Neural Direta com <i>Backpropagation</i>	Valor de R ²	Nenhum	Previsão superior para rede com memória de curto-tempo I - Delay
Condutividade Hidráulica da terra. (TAMARI et al., 1996)	-	Rede Neural RBF	MSE	Regressão Multilinear	A rede obteve melhores resultados que o modelo de regressão quando as incertezas eram pequenas.

Série de Tempo	Memória de Curto-Tempo	Componente Preditor	Medidas de Erro	Métodos de Comparação	Resultados Obtidos
Cinco séries de tempo (1 de Hénon e 4 do tipo AR(2)) (STERN, 1996)	I - Delay	Rede Neural Direta com Backpropagation	RMSE	Outros métodos de previsão não-lineares	O desempenho da RNA foi melhor para séries com pouco ou nenhum componente estocástico.
Índice mensal de passageiros das linhas aéreas americanas. (FARAWAY & CHATFIELD, 1998)	I - Delay	Rede Neural Direta com Backpropagation	MSE	Box & Jenkins e Método de Holt-Winters	Concluíram que para obter bons resultados com uma rede, habilidades de modelagem tradicionais devem ser usadas na seleção do modelo
Cotação diária de moeda (ANDRADE et al., 1997)	I - Delay (Distribuição aleatória)	Rede Neural Direta com Rprop (Resilient Backpropagation)	MSE MPE	Box & Jenkins	Box & Jenkins obteve melhores resultados do que a RNA.
Previsão de vazões (KADOWAKI et al., 1997) e (VALENÇA & VASCONCELOS, 1997) (LIMA Jr. & DE CARVALHO, 1997)	I - Delay	Rede Neural Direta com Backpropagation	MPE MAE MAPE	Box & Jenkins Regressão Múltipla	Redes neurais obtiveram os melhores resultados
Séries Padrões (TANG & FISHWICK, 1991)	I - Delay O - Delay	Rede Neural Direta com Backpropagation	MAPE	Box & Jenkins	Para séries longas, superioridade para a metodologia Box & Jenkins na previsões de curto prazo, as RNAs são melhores para longo prazo. Para séries curtas, RNAs são superiores em ambos os casos.

Série de Tempo	Memória de Curto-Tempo	Componente Preditor	Medidas de Erro	Métodos de Comparação	Resultados Obtidos
Preço de ações (MUELLER, 1996)	I - Delay	Rede Neural Direta com Backpropagation	MAPE	Méida Móvel Alisamento Exponencial Simples	RNAs obtiveram os melhores resultados
Carga ativa nodal em sistemas de transmissão e distribuição de energia elétrica (DE OLIVEIRA, 1999)	-	Rede Neural RBF	MAPE EPmax CV	Rede Neural Direta com Backpropagation	O modelo MLP produziu previsões mais precisas.
Preço do ouro no mercado internacional. (ABELÉM, 1994)	TI - Delay TO - Delay	Rede Neural Direta com Backpropagation	Coef. U de Theil MSE NRMSE POCID	Box & Jenkins	Coef. U de Theil para a RNA (1,0028), para o modelo Box & Jenkins (1,0018)
Série A (WAN, 1994)	TI - Delay	Rede Neural Direta com ackpropagation temporal	NMSE	Competição - Santa Fe Institute	RNA melhor método de previsão da competição
Série C, D e F (ZHANG & HUTCHINSON, 1994)	I - Delay	Rede Neural Direta com Backpropagation temporal			
Série C (MOZER, 1994)	I - Delay TIS - 0 Híbrido TIS - 0 e I - Delay	Rede Neural Direta Com Backpropagation			

4.6 CONSIDERAÇÕES FINAIS

O problema de previsão pode ser considerado um caso especial dos problemas de aproximação de funções, onde os valores da função são representados por uma série temporal.

De acordo com CYBENKO (1988), apud BRAGA et al. (2000), uma rede com uma camada intermediária pode implementar qualquer função contínua. A utilização de duas camadas intermediárias permite a aproximação de qualquer função (CYBENKO (1988), apud BRAGA et al. (2000)). Estes resultados teóricos fazem das RNA diretas boas candidatas para o mapeamento de qualquer função complexa, não-linear (BARRETO, 1997), que pode governar mudanças em uma série temporal.

Porém, a previsão envolve o processamento de padrões que se desenvolvem com o passar do tempo - a resposta apropriada em um ponto particular do tempo depende não apenas do valor corrente da observação, mas também dos valores prévios desta.

Para uma rede neural ser considerada dinâmica, é preciso que possua memória. Esta pode ser provida pela introdução de atraso no tempo, ou pela utilização de redes recorrentes.

MOZER, em 1994, criou uma taxonomia para RNAs para processamento temporal. Segundo esta taxonomia, uma RNA consiste em dois elementos: uma memória de curto-tempo (Pré-processador) e um componente preditor (uma rede neural direta MLP).

A memória calcula o estado da rede e o componente preditor usa este estado para calcular a saída da rede.

A taxonomia de MOZER aborda a memória de curto-tempo, mas, a respeito do componente preditor, só define que este deve ser uma rede neural direta.

Uma vez que o desempenho obtido por uma RNA, principalmente de redes MLP, depende da topologia da rede (tamanho, estrutura e conexões) e dos parâmetros do algoritmo de treinamento (taxa de aprendizado, termo de momento e outros), a determinação de uma rede adequada ainda não é uma tarefa simples.

CAPÍTULO 5 - MODELAGEM DE REDES NEURAIS ARTIFICIAIS PARA PREVISÃO DE SÉRIES TEMPORAIS

Identificar os fatores determinantes no desempenho das RNAs e os parâmetros que controlam estes fatores;

5.1 CONSIDERAÇÕES NO PROJETO DE REDES NEURAIS ARTIFICIAIS

O pouco entendimento das complexas inter-relações entre os parâmetros que influenciam no desempenho da rede, como função de ativação, função erro, parâmetros de controle do gradiente descendente e do tempo de treinamento e arquitetura da rede, permite um grau de liberdade extra no projeto de RNAs. Este grau de liberdade na escolha dos parâmetros adequados, torna as aplicações altamente dependente da experiência do projetista e da experimentação manual para que se obtenha sucesso.

5.1.1. FATORES DETERMINANTES NO DESEMPENHO DAS RNAs

De acordo com ABELÉM (1994) e MUELLER (1996), o desempenho das RNAs em geral é avaliado em função de três fatores:

- Convergência
- Generalização
- Escalabilidade

A convergência é um fator importantíssimo para o bom desempenho das RNAs, pois diz respeito à capacidade das RNAs de aprender um determinado conjunto de dados definido no procedimento de treinamento. A convergência é o que indica quando o procedimento de treinamento é capaz de aprender a classificar corretamente os padrões apresentados, sob que condições e sob quais exigências computacionais. RNAs com topologias fixas que utilizam o algoritmo de retropropagação provam a convergência mostrando que em um tempo limite, quando o tempo de treinamento tende ao infinito, o erro minimizado pelo método do gradiente descendente tende a zero. Outros métodos (sub-ótimos) provam a convergência mostrando que quando o tempo de treinamento

tende ao infinito, o método pode classificar o número máximo de possíveis mapeamentos com uma probabilidade arbitrariamente grande $p \leq 1$. A utilização da convergência como fator determinante para o desempenho da rede é feita por motivos óbvios, já que se uma determinada RNA não consegue aprender o conjunto treinamento que lhe foi apresentado (não converge), dificilmente conseguirá oferecer bons resultados no momento dos testes.

A generalização diz respeito à capacidade de a rede aprender através de um conjunto reduzido de exemplos e posteriormente dar respostas coerentes para dados não-conhecidos. Apesar das várias aplicações bem sucedidas de RNAs descritas na literatura (CHAKRABORTY et al., 1992), (PODDIG, 1993), (MUELLER, 1996), (DE OLIVEIRA, 1999), e outros, as condições para que boas generalizações sejam encontradas ainda não são totalmente conhecidas. Estudos teóricos citados em REFENES (1992) têm mostrado que a probabilidade de generalizações corretas depende do tamanho do espaço de hipóteses (ex. número total de redes consideradas), do tamanho do espaço de soluções (ex. conjunto de RNAs que fornece boas soluções) e o número de exemplos do treinamento. Se o espaço de hipóteses é muito grande e/ou o número de exemplos do treinamento é muito pequeno, então existirá um vasto número de RNAs, as quais são consistentes para o conjunto de treinamento. Porém, somente uma pequena porção deste conjunto se encontra no espaço de soluções corretas, fazendo com que seja provável uma generalização pobre.

De uma forma geral, a diminuição dos parâmetros livres (número de conexões) da rede aumenta a probabilidade de uma boa generalização. Contudo, isto deve ser feito sem reduzir demais o tamanho da rede, já que esta redução pode implicar na incapacidade da rede em computar a função desejada. Outro ponto importante é a escolha de um conjunto de treinamento bastante representativo, apresentando as relações mais importantes que devem ser aprendidas pelas RNAs. A grande maioria dos pesquisadores experientes concorda que uma boa representação dos dados de entrada é um ponto de fundamental importância para o sucesso de uma aplicação.

Por fim, a escalabilidade, que diz respeito a ambos, convergência e generalização, é relativa ao procedimento de ajustar os parâmetros das RNAs para que elas consigam convergir e generalizar bem. Sabe-se que para uma RNA ter um bom desempenho, ela

deve ser grande o suficiente para aprender o problema (convergir), mas também pequena o bastante para generalizar bem.

5.1.2. CONTROLE DOS FATORES DETERMINANTES NO DESEMPENHO DAS RNAs

O controle dos fatores determinantes de desempenho se dá pela escolha adequada dos parâmetros da rede como: função de ativação, função erro, parâmetros de controle do gradiente descendente e do tempo de treinamento e arquitetura da rede, os quais, como dito anteriormente, variam de aplicação para aplicação e não possuem regras de escolha bem definidas.

5.1.2.1. PADRONIZAÇÃO

Apesar de não controlar diretamente os fatores determinantes de desempenho das RNAs, a padronização pode, freqüentemente, melhorar a performance das RNAs através da remoção de características insignificantes do conjunto de dados. As informações relevantes das séries são, geralmente, independentes de valores de compensação ('offset') e de desvios padrões. Estas características podem obscurecer a tarefa das RNAs. A padronização, entretanto, admite várias formas e exige cuidados, já que é muito fácil remover, inadvertidamente, informações preciosas dos dados. Por outro lado, a padronização correta dos dados pode melhorar consideravelmente o desempenho de uma rede neural (MAKRIDAKIS et al., 1998).

REMOÇÃO DA MÉDIA

Em muitos casos, a informação significativa de uma série está contida em pequenas variações que se encontram adicionadas a um grande valor de compensação¹. Este valor de compensação pode ser retirado se é sabido que este é irrelevante para o sistema e está inibindo o treinamento. Para isso, basta que se retire de todos os pontos da série a média da mesma. Assim, de acordo com ABELÉM (1994), para cada componente dos valores da série, tem-se:

$$W_i = Y_i - \bar{Y} \text{ onde, } \bar{Y} = \sum Y / n \quad (55)$$

¹ Grandes valores de compensação podem levar os neurônios não lineares à saturação e, assim, inibir o treinamento.

onde:

- W_t → Observação transformada no tempo t
 Y_t → Observação original no tempo t
 \bar{Y} → Média das observações originais
 n → Número de observações da série Y

DIFERENCIAÇÃO

Outro tipo de padronização é denominada diferenciação dos dados. Esta padronização é utilizada principalmente quando se tem valores com pequenas variações absolutas, ficando a informação significativa na variação percentual dos mesmos. As pequenas variações em termos absolutos comumente camuflam a informação importante, fazendo com que a rede tenha um desempenho ruim. Desta forma, conforme ABELÉM (1994), para cada componente dos vetores da série, tem-se:

$$W_t = \frac{Y_t - Y_{t-1}}{\Delta_t} \quad \text{onde, } \Delta_t = (t - (t-1)) \quad (56)$$

PADRONIZAÇÃO DA MAGNITUDE

Grandes variações na magnitude dos componentes do vetor é outro fator que pode confundir a rede. Nestes casos, faz-se a padronização nas magnitudes do vetor, dividindo cada componente do vetor pelo desvio padrão calculado sobre o total de componentes. Com isto, segundo ABELÉM (1994), para cada componente dos vetores da série, tem-se:

$$W = \frac{Y}{S} \quad \text{onde, } S = \sqrt{\frac{\sum (Y - \bar{Y})^2}{n-1}} \quad (57)$$

TRANSFORMAÇÃO MATEMÁTICA

A transformação matemática dos dados, como o logaritmo, é usada principalmente em casos onde a magnitude dos dados varia de muitas ordens de grandeza, mas a informação significativa está contida na variação percentual. Por exemplo, quando os maiores valores dos dados excedem a região da rede, as variações percentuais nos pequenos valores provavelmente serão perdidas. Neste caso, o logaritmo dos dados irá ajustar as faixas, tal que os maiores valores serão comprimidos mais que os

pequenos, alocando uma faixa constante para os desvios percentuais, independente do valor dos dados. Alguns exemplos de padronização não linear para cada componente dos valores da série, podem ser encontrados no Quadro 5.1.

Quadro 5.1 - Exemplos de padronização não linear.

raiz quadrada	$W_t = \sqrt{Y_t}$	↓
raiz cúbica	$W_t = \sqrt[3]{Y_t}$	Aumenta
logaritmo	$W_t = \log(Y_t)$	a
inverso	$W_t = -1/Y_t$	força
		↓

Fonte: MAKRIDAKIS et al. (1998).

AJUSTES DE CALENDÁRIO

Alguma das variações em uma série temporal ocorrem devido à variação no número de dias de cada mês.

A duração do mês pode ter um efeito sazonal grande, uma vez que um mês pode diferir por aproximadamente $\frac{31-28}{30} = 10\%$. Se esta não for removida, aparecerá como um efeito sazonal, dificultando a interpretação da série. Os dados são ajustados facilmente por MAKRIDAKIS et al.(1998), com o uso da Equação 58:

$$W_t = \frac{365,25/12}{n^{\circ} \text{ de dias mês } t}$$

(58)

Para todos estes tipos de padronização, quando a característica está sobre todos os componentes dos vetores, faz-se uma padronização total. Caso os componentes dos vetores representem diferentes tipos de dados, faz-se uma padronização vertical para cada tipo. Já no caso em que componentes adjacentes do vetor representam amostras consecutivas de um processo contínuo, usa-se a padronização horizontal. Por fim, em casos onde os dados possuem uma distribuição multi-modal, com os vários modos tendo diferentes médias e desvios padrões, usa-se a padronização local (ABELÉM, 1994).

Outras formas de padronização podem ser encontradas em (MAKRIDAKIS et al., 1998).

5.1.2.2. FUNÇÕES DE SAÍDA OU TRANSFERÊNCIA

Existem vários tipos de funções utilizadas, como funções lineares, de limiar ou funções não-lineares, que simulam mais precisamente as características dos neurônios biológicos.

A escolha da função de saída varia muito em função da aplicação e do algoritmo de treinamento utilizado. As mais comuns funções de saída usadas em neurônios individuais são a função linear, a função logística (assimétrica) e a função tangente hiperbólica (simétrica), descritas abaixo. Todavia, segundo (DE AZEVEDO, 1999), qualquer função monotonicamente crescente e contínua tal que $x \in \mathcal{R}$ e $y(x) \in [-1, 1]$, pode ser usada.

- a função linear:

$$y(x) = x \quad (59)$$

- a função logística, que é a mais popular função unipolar:

$$y(x) = \frac{1}{1 + e^{-kx}} \quad (60)$$

- e a função tangente hiperbólica, que é a mais popular função bipolar:

$$y(x) = \tanh(kx) = \frac{e^{kx} - e^{-kx}}{e^{kx} + e^{-kx}} \quad (61)$$

onde k é um escalar positivo.

Outras funções podem ser derivadas pelo ajuste da constante k . Quanto mais k cresce, mais abrupta a função é nas proximidades de $x = 0$.

Essas funções são, também, denominadas de "com características sigmoidais", e quando $k \rightarrow \infty$ elas se tornam "*hard-limiting functions*", as quais descrevem modelos de neurônios discretos (DE AZEVEDO, 1999).

Em aplicações que possuem valores altos no conjunto de treinamento, utiliza-se a função não-linear somente nos neurônios das camadas intermediárias, ficando os neurônios da camada de saída com funções lineares. Este artifício tem o objetivo de evitar a "paralisia" da rede, pois como os valores de entrada são altos, mesmo pequenos erros podem ser grandes em valores absolutos, forçando os neurônios a trabalharem em regiões onde a derivada das funções não-lineares é praticamente nula. No caso do treinamento com retropropagação, como o erro que é retropropagado é diretamente proporcional a esta derivada, o processo de treinamento fica paralisado e a rede não converge (KOVÁCS, 1996).

A escolha entre as funções simétricas ou assimétricas depende muito da aplicação. Em geral, quando se trabalha com números negativos, opta-se por funções simétricas como a *tanh*, já que estas são bipolares.

5.1.2.3. FUNÇÃO DE ERRO

A função de erro informa o grau de convergência das RNAs. Através dela, é possível saber o quanto as RNAs estão aprendendo os padrões apresentados durante o treinamento. A função de erro mais comumente usada é a MSE dada pela Equação 3, descrita no Quadro 2.1, pg. 13.

Apesar de nem sempre ser a função mais indicada, devido aos problemas causados por trabalhar com as médias, a MSE é a função mais utilizada pela sua facilidade de cálculo e por ser uma métrica encontrada em praticamente todos os pacotes computacionais que trabalham com estatística e RNAs, facilitando assim a comparação dos resultados.

Uma outra opção de função de erro é a MAD (Desvio Absoluto Médio), a qual é dada pela fórmula:

$$MAD = \frac{1}{n} \sum_{i=1}^n |Y_i - F_i| \quad (62)$$

A diferença básica entre as duas medidas é que a primeira eleva o erro ao quadrado, penalizando erros maiores com muito mais rigor.

5.1.2.4. PARÂMETROS DE CONTROLE DO GRADIENTE DESCENDENTE E DO TEMPO DE TREINAMENTO

O terceiro importante mecanismo para controlar o desempenho do aprendizado das RNAs é a escolha dos parâmetros de controle do gradiente descendente e do tempo de treinamento. Pesquisadores têm experimentado termos extras na fórmula do gradiente descendente como o termo de momento, segunda derivada, entre outros (BRAGA et al., 2000). O termo de momento consiste em incluir no cálculo de troca de pesos um termo proporcional de troca dos pesos da iteração imediatamente anterior (LOESCH & SARI, 1996). A inclusão do termo de momento na fórmula de ajuste dos pesos aumenta a velocidade de aprendizado (aceleração), reduzindo o perigo de instabilidade (ABELÉM, 1994).

A utilização da segunda derivada é um artifício que tem sido usado por vários métodos de treinamento e que chega a reduzir por um fator de cem, o tempo de treinamento (ABELÉM, 1994). Nesta técnica, ao invés de utilizar a inclinação da superfície do erro (primeira derivada) como faz o algoritmo de retropropagação, utiliza-se a taxa de variação da inclinação (segunda derivada) desta superfície.

Outro termo de fundamental importância no controle do gradiente descendente e no tempo de treinamento é a taxa de aprendizagem (η). Esta essencialmente controla o grau segundo o qual o gradiente afeta os pesos. Após a propagação dos erros derivativos quadráticos, atualiza-se o vetor dos pesos das conexões na entrada de cada neurônio, adicionando-se a este uma parcela do produto do erro derivativo quadrático associado pelo vetor de entradas. Seu valor, segundo LOESCH & SARI (1996), influencia na velocidade de convergência do processo, assim como na sua estabilidade. Se for relativamente grande, tende a acelerar o processo de convergência pois força a passos maiores a cada iteração, mas, por outro lado, pode produzir oscilações em torno do ponto ótimo, isto é, não permite a estabilização. Uma pequena taxa estabiliza o processo, mas resulta em uma aprendizagem lenta, e aumenta sensivelmente a propensão para mínimos locais.

De acordo com ABELÉM (1994), uma estratégia que tem se mostrado bastante eficiente é o uso de taxas de aprendizado dinâmicas, as quais começam com valores altos e vão diminuindo à medida que o treinamento vai evoluindo.

5.1.2.5. ARQUITETURA DA RNA

A arquitetura da rede é muito importante para as capacidades de convergência e generalização das RNAs. A escolha de uma arquitetura adequada, nem tão grande e nem tão pequena, é a principal responsável pelo sucesso de uma aplicação. Para se obter uma boa habilidade de generalização, deve-se projetar a rede colocando dentro dela o máximo de conhecimento possível sobre o problema (ex. a topologia do espaço de entradas) e limitar o número de conexões adequadamente.

O número de neurônios da camada de entrada e de saída é determinado pelo problema em si. A dificuldade está na determinação do número de neurônios da camada intermediária. Este, na maioria das vezes, é determinado por heurísticas ou *regras ad-hoc* como a de LOESCH & SARI (1996), que diz: *"o número de neurônios da camada intermediária é igual à média geométrica do número de neurônios das camadas de entrada e saída"*. Ou ainda, segundo EBERHART apud BRASIL (1999), *"o número de neurônios da camada intermediária é igual a raiz quadrada da quantidade de neurônios na camada de entrada somado com o número de neurônios da camada de saída"*. O uso dessas heurísticas ou outra, às vezes funciona para certos tipos de problemas.

Existem várias técnicas que tentam encontrar a arquitetura ideal através da diminuição de neurônios redundantes e do compartilhamento de pesos (BRAGA et al., 2000). Porém, nenhuma conseguiu resultados convincentes para todas as aplicações. Existe também a possibilidade de fazer uma busca no espaço de todas as arquiteturas possíveis. Para isso, deve-se treinar cada uma das arquiteturas separadamente e então avaliá-las com uma função de erro apropriada que incorpore o desempenho e o número de unidades. A busca também pode ser feita através de um algoritmo genético, tal que as boas arquiteturas encontradas em uma triagem sejam mantidas para serem combinadas com boas arquiteturas de outras triagens (BRAGA et al., 2000). Contudo, estes tipos de busca se mostram pouco práticos para aplicações que utilizam grandes RNAs, onde cada

treinamento exige esforço computacional. A verdade é que a arquitetura ideal das RNAs ainda continua sendo determinada principalmente pela experimentação manual e pela experiência e sensibilidade do projetista.

5.2 CONSIDERAÇÕES FINAIS

O Quadro 5.2 sumariza os principais fatores determinantes no desempenho das RNAs e os vários parâmetros que permitem o controle destes fatores.

Quadro 5.2 - Principais fatores determinantes no desempenho das RNAs.

Parâmetros de controle		Aplicação/Influência dos Parâmetros
	<i>Padronização</i>	
	→ Remoção da média	séries com pequenas variações adicionadas a um grande valor de compensação.
	→ Diferenciação	séries com pequenas variações absolutas.
	→ Norm. magnitude	séries com grandes variações de magnitude.
	→ Padr. Matemática	séries com variações de magnitude de muitas ordens de grandeza.
Convergência	<i>Funções de ativação</i>	
	→ Sigmóide	para padrões de treinamento unipolares.
	→ Tangente hiperbólica	para padrões de treinamento bipolares.
	→ Sigmóide/linear	para padrões de treinamento unipolares e muito grandes.
	→ Tangente/linear	para padrões de treinamento bipolares e muito grandes.
Generalização	<i>Funções de erro</i>	
	→ MAD	trata todos os erros igualmente.
	→ MSE	penaliza os erros maiores com mais rigor.
Escalabilidade	<i>Tempo de treinamento e gradiente descendente</i>	
	→ Taxa de aprendizado	influencia a magnitude das mudanças dos pesos.
	→ Termo de momento	treinamento mais rápido, com menor risco de oscilações e de mínimos locais.
	→ Segunda derivada	diminui o tempo de treinamento.
	<i>Arquitetura da rede</i>	
	→ Técnicas de diminuição de neurônios redundantes	pouco eficiente.
	→ Algoritmos genéticos	problemas para RNAs muito grandes.
	→ Experimentação	a mais utilizada.

CAPÍTULO 6 - ESTUDO DE CASO

*Analisar o comportamento da série de preço de ovos;
aplicar a taxinomia de MOZER ao problema;
avaliar os resultados obtidos.*

O objetivo do trabalho é propor um modelo que permita a previsão do preço futuro do ovo, pago ao produtor, para um período de no mínimo 318 dias (10,6 meses) à frente.

O modelo é fundamentado na técnica de RNA para processamento temporal. Para resolução do problema proposto, utilizou-se uma rede neural direta multicamadas com uma memória de curto-tempo do tipo *TO-Delay* estática. A validação do modelo foi feita através da comparação com os tradicionais modelos ARIMA, muito utilizados para previsão de séries temporais econômicas.

6.1 DADOS

No estudo de caso, as RNAs foram modeladas para fazer previsões do preço do ovo branco (tipo grande) pagos ao produtor. Os dados utilizados provêm de levantamentos efetivados pelo Instituto de Economia Agrícola - IEA, referindo-se a médias mensais para os preços reais (deflacionados pelo IGP - DI), no período de Janeiro/85 a Dezembro/99 (Fig. 6.1). Os dados utilizados se encontram no Anexo B.1.

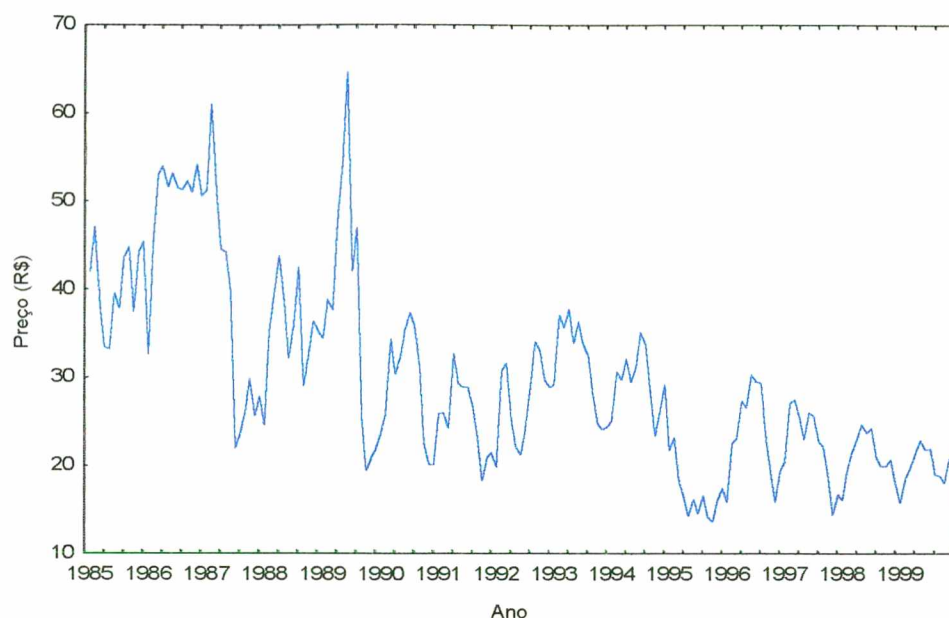


Figura 6.1 - Preço real do ovo branco (tipo grande) pago ao produtor de São Paulo: Jan/85 a Dez/99.

Os ovos brancos, tipo grande, segundo o decreto nº. 56585, de 20/07/65, são ovos com peso de no mínimo 55g. por unidade, e no máximo 660g. por dúzia. Segundo especialistas na área, é o tipo de ovo mais produzido e consumido no Brasil. Tentou-se assim, abranger a maior parte da produção.

Todavia, de acordo com MARTINS (1996), o comportamento do preço dos vários tipos de ovos (extra, grande, médio e pequeno), como também, nos vários mercados (preço ao produtor, no atacado e varejo) é quase constante, ou seja, sobem ou caem juntos, na mesma proporção. Assim sendo, o modelo estimado aqui, poderá ser usado para previsão do preço dos diversos tipos de ovos, e nos diversos mercados.

A série de preços fornecida pelo IEA refere-se ao estado de São Paulo, o que não deve impedir que as conclusões obtidas se apliquem, em termos gerais, à avicultura brasileira, já que São Paulo é responsável por 40,07% da produção nacional (MARTINS, 1996).

Inicialmente, os dados foram padronizados aplicando-se o logaritmo e ajustando estes conforme o calendário, como descrito no item 5.1.2.1 (Fig. 6.2).

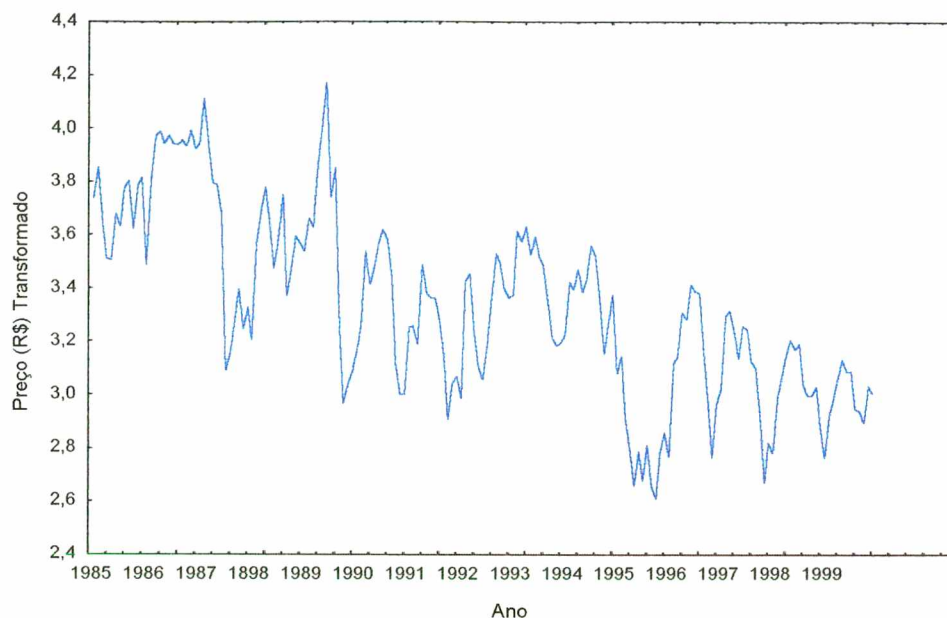


Figura 6.2 - Preço real do ovo branco (tipo grande) pago ao produtor de São Paulo, após a transformação.

Após, o conjunto de dados foi dividido em duas partes:

- a primeira, relativa ao período de Janeiro/85 a Dezembro/97, foi utilizada para o treinamento das RNAs e identificação do modelo ARIMA (Fig. 6.3);
- a segunda parte, utilizada para teste, corresponde aos dois anos restantes, Janeiro/98 a Dezembro/99 (Fig. 6.4).

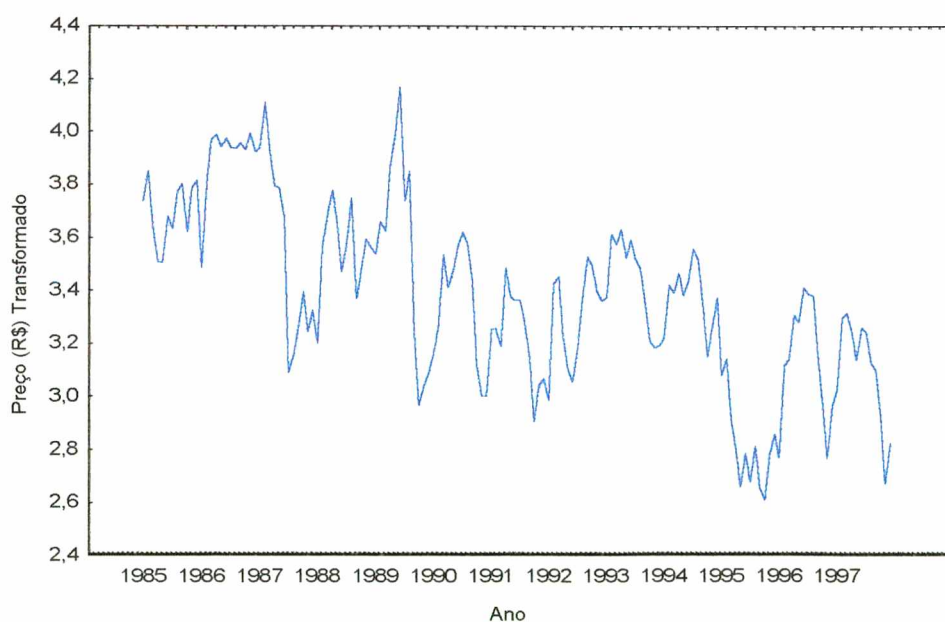


Figura 6.3 - Conjunto de treinamento utilizado. Dados transformados de Jan/85 a Dez/97.

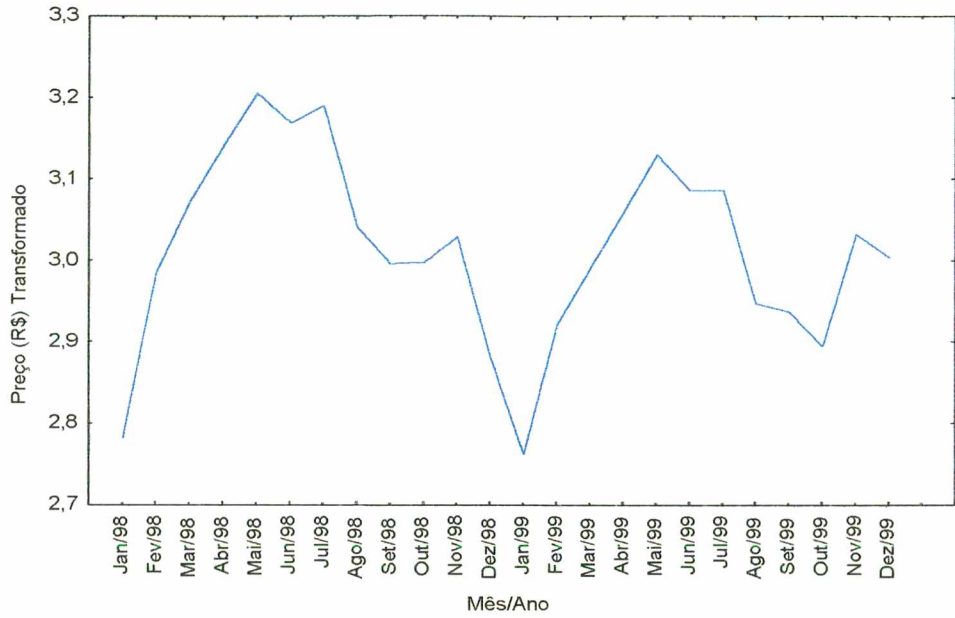


Figura 6.4 - Conjunto de teste utilizado. Dados transformados de Jan/98 a Dez/99.

Para o conjunto de treinamento da RNA, os dados foram ainda normalizados entre [-1 e 1], conforme a Equação 63.

$$V_{norm} = \frac{(2 \times V - (V_{max} + V_{min}))}{(V_{max} - V_{min})} \tag{63}$$

onde:

- V_{norm} → valor normalizado
- V → valor original
- V_{min} → valor mínimo
- V_{max} → valor máximo

Na RNA, a normalização dos dados se faz necessária devido a função de ativação utilizada na rede (Função tangente hiperbólica), que é bipolar.

6.2 CARACTERÍSTICAS NÃO-OBSERVÁVEIS DA SÉRIE

Utilizando-se um método clássico de decomposição (*Census II¹*), é possível decompor a série em seus componentes básicos não-observáveis (tendência-ciclo, sazonalidade e ruído aleatório). Uma vez conhecidas estas características básicas da série, é possível criar um ou mais neurônios que busquem representar estas características da série. Assim sendo, pode-se usar a modelagem econométrica de

componentes não-observáveis como uma fase preliminar na montagem da memória de curto-tempo da RNA.

Como podemos observar na (Fig. 6.1) , o preço do ovo, no período de Jan/85 a Dez/99, apresentou sazonalidade, ciclo e uma nítida tendência de queda como características marcantes.

O pico de preço no período ocorreu em 1989, com repiques decrescentes em 1993, 1990, 1994 e 1992. O preço mínimo do ovo verificou-se em 1995.

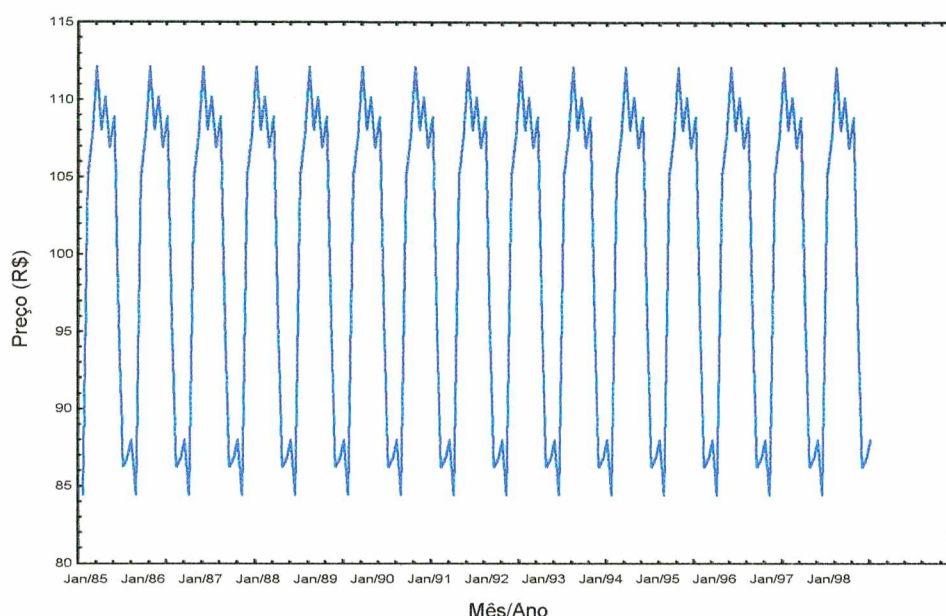


Figura 6.5 - Componente sazonal da série.

O fator sazonal da série, Fig. 6.5, mostra três períodos de pico anual no preço do ovo pago ao produtor, durante os meses de março, abril, junho, julho e agosto, em função do outono e inverno (onde, o preço do ovo aumenta devido a produção menor - característica do ciclo biológico das aves). Além disso, pode-se perceber um período de redução sazonal nos meses de janeiro, outubro e novembro, onde a produção é maior, aumentando a disponibilidade do produto e ocasionando a queda nos preços.

¹ Uma descrição detalhada sobre os métodos de decomposição de séries temporais pode ser encontrada em MAKRIDAKIS et al. (1998).

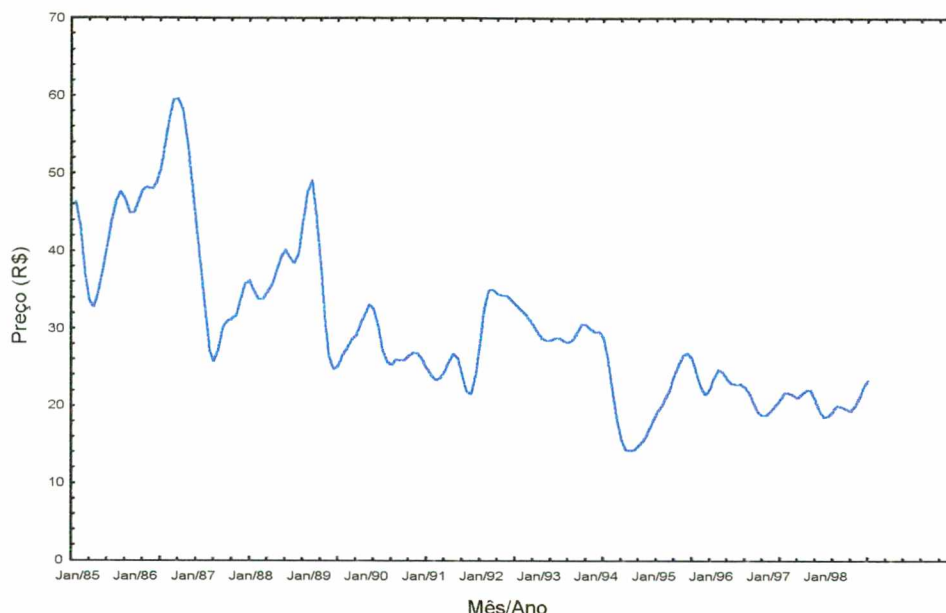


Figura 6.6 - Componente tendência/ciclo da série

A componente tendência-ciclo, apresentada na Fig. 6.6, aponta para um declínio não linear da série, indicando, portanto, uma redução do preço do ovo ao longo do período analisado.

Essa nítida tendência de queda, segundo MAKRIDAKIS et al. (1998), é normal para a maioria dos produtos e serviços padronizados. Excluindo a inflação, o preço desses produtos diminui com o tempo, devido ao aumento da oferta (pela adoção de novas tecnologias, melhoramento genético, entre outros fatores) sobre a demanda. De acordo com o autor, com a revolução da informação, o declive tenderá a acelerar ainda mais.

Dessa forma, essa tendência, pode ser extrapolada com um razoável grau de certeza, uma vez que esta persevera há muito tempo, sendo o resultado de um sistema econômico de livre competição.

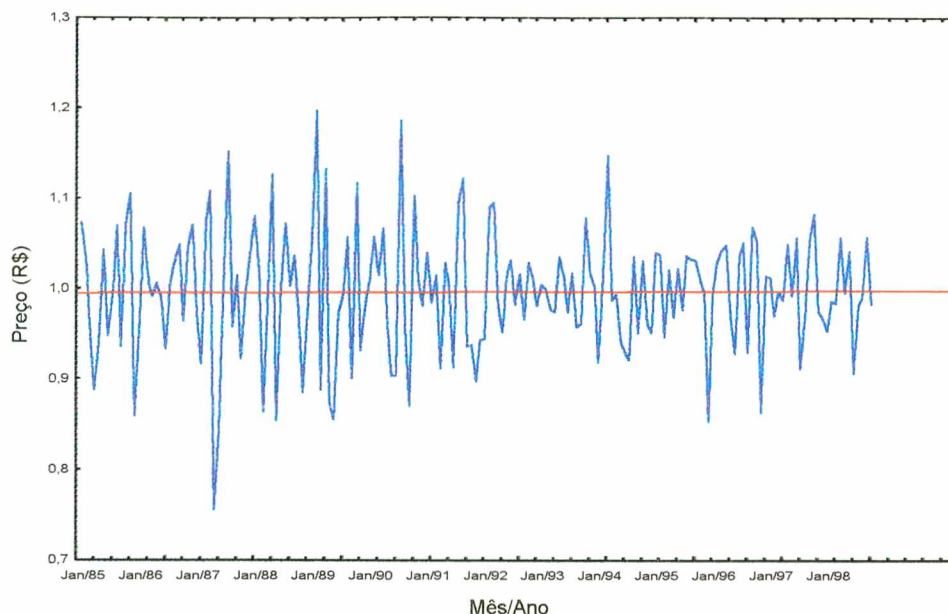


Figura 6.7 - Componente irregular da série.

Na Fig. 6.7, é apresentado o componente irregular da série, que apresenta tudo aquilo que não pode ser explicado estatisticamente durante a análise da série, ou seja trata-se do componente aleatório da série. Como pode-se observar, há uma redução na variância do componente irregular para o período após 1991.

A simples inspeção do gráfico mostrado na Fig. 6.7 indica a dificuldade enfrentada para a realização de prognósticos que apresentem erros suficientemente pequenos.

6.3 MODELOS PARA PREVISÃO DE PREÇOS DE OVOS

6.3.1. RNA

Seguindo a taxinomia de MOZER (1994), o modelo de rede proposto é composto por:

- uma memória de curto-tempo *TO-Delay* Estática e
- uma RNA direta MLP, totalmente conectada, com treze neurônios na camada de entrada, dez na camada intermediária e um na camada de saída, que pode ser representada por RNA(13, 10, 1) (Fig. 6.8).

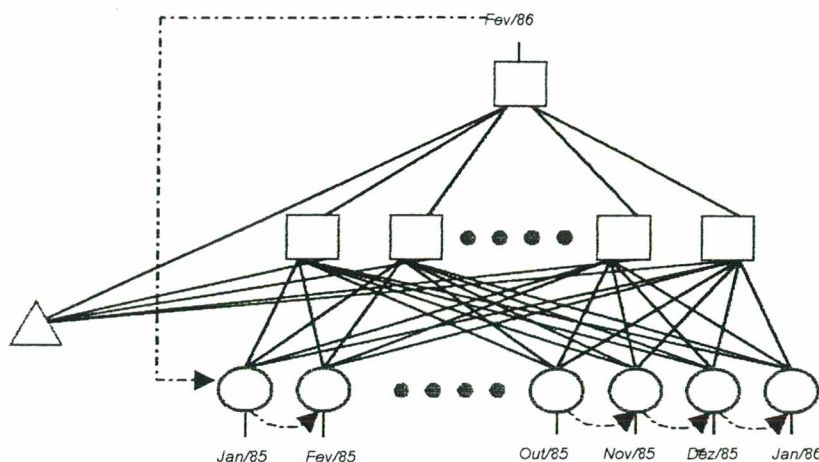


Figura 6.8 - Arquitetura de RNA proposta para a previsão de preço de ovos.

A função de ativação utilizada foi a mesma para todos os neurônios, ou seja, a função bipolar tangente hiperbólica. O treinamento foi feito utilizando-se o método padrão de retropropagação do erro com momento, com a seguinte configuração:

- Inicialização aleatória dos pesos;
- o conjunto de treinamento foi apresentado 200 000 vezes à rede;
- constante de aprendizado (α) = 0,05;
- termo de momento (β) = 0,4 durante todo o treinamento;
- função de erro: MSE (Erro Quadrático Médio);
- erro máximo admissível: Erro Quadrático Médio < 1e-5.

O conjunto de treinamento (que pode ser visto no Anexo B.2) e teste foram pré-processados (como descrito anteriormente) e o último, pós-processado no final do treinamento, pela inversa das funções utilizadas.

DE OLIVEIRA (1999) em sua dissertação de mestrado, comparou o desempenho de uma rede neural direta MLP com uma rede RBF, para previsão automática de carga ativa nodal de curto prazo em sistemas de transmissão. Em seu trabalho, DE OLIVEIRA concluiu que o modelo de rede RBF além de convergir mais rápido para uma solução ótima é mais fácil de ajustar que um modelo MLP. Porém, em termos de precisão de resultados e capacidade de generalização, ele afirma que o modelo MLP supera os modelos RBF.

Como para o problema de previsão de preços (a longo prazo) o fator mais importante é a precisão dos resultados e a capacidade de generalização, o modelo MLP é mais adequado.

6.3.2. MODELO SARIMA

Para avaliar o modelo proposto, este foi comparado com o método estatístico tradicional para previsão de séries temporais, Box & Jenkins.

Optou-se, neste trabalho, pelo uso do modelo auto-regressivo, integrado, de média móvel sazonal $SARIMA(p, d, q)(P, D, Q)_s$, por se tratar do preço de um produto agrícola que recebe a influência de fatores sazonais. A estimativa seguiu as etapas de identificação, estimação dos parâmetros, verificação e previsão, descritas no Item 3.3.

No presente trabalho, utilizando-se do Sistema Computacional para Análise Estatística de Dados - STATISTICATM, versão 5.0, realizou-se o processo de identificação (através das funções de autocorrelação e autocorrelação parcial, que podem ser encontradas no Anexo A), selecionando-se uma série de modelos possíveis a serem testados. O modelo que apresentou o melhor desempenho foi modelo $SARIMA(0, 1, 0)(2, 1, 0)_{12}$.

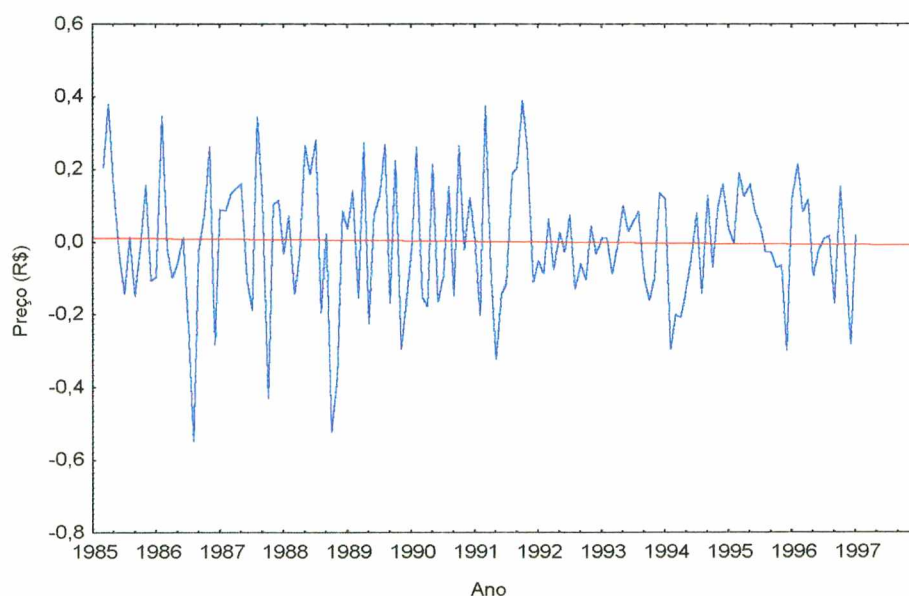


Figura 6.9 - Resíduos do modelo $SARIMA(0,1,0)(2,1,0)_{12}$

A estimação dos parâmetros foi feita pelo método de máxima verossimilhança, que segundo FAVA (2000b) é superior ao método de mínimos quadrados.

Tabela 6.1 - Parâmetros estimados para o modelo (0, 1, 0)(2, 1, 0)₁₂

Parâmetro	Valor Estimado	Erro Padrão	t	p	Intervalo de confiança (95%)
P _s (1)	-0,647	0,085	-7,569	0,000	[-0,815, -0,478]
P _s (2)	-0,193	0,089	-2,169	0,032	[-0,369, -0,017]

Como pode-se observar na Tabela 6.1, os dois parametros estimados são significativamente diferentes de zero (valores de *t* alto (7,569 e 2,169)). O valor de *p* indica a probabilidade de se obter um valor de $|t|$ tão extremo quanto o valor calculado, se de fato o parâmetro for igual a zero. Assim, nesse caso, como o valor de *p* para os dois parâmetros foi menor que 0,05, podemos concluir que os dois parâmetros são muito significativos.

A Fig. 6.9 mostra os resíduos após a aplicação do modelo SARIMA (0, 1, 0)(2, 1, 0)₁₂. Como pode-se notar, os resíduos apresentam valores sempre bastante próximos de zero.

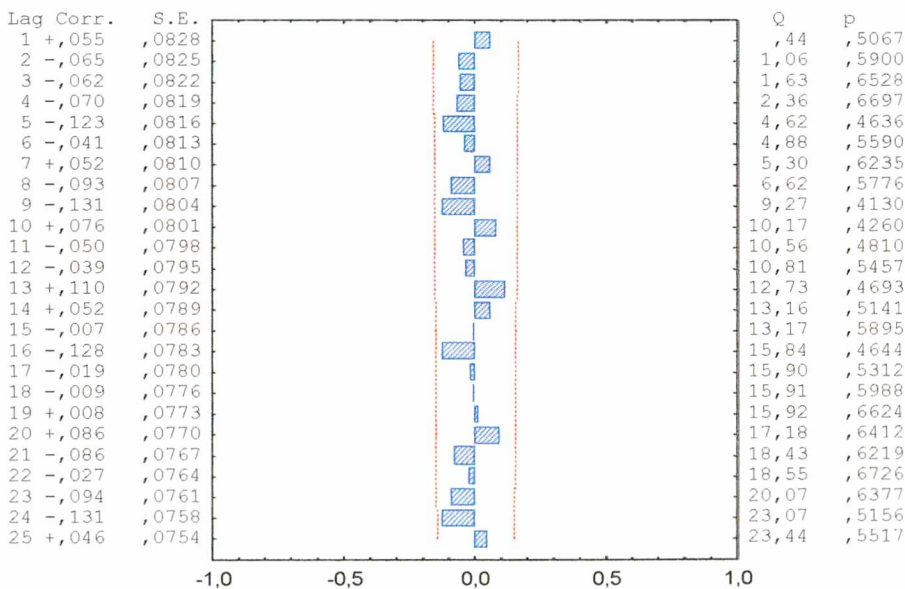


Figura 6.10 - Função de autocorrelação (FAC) dos erros do modelo SARIMA (0,1,0)(2,1,0)₁₂.

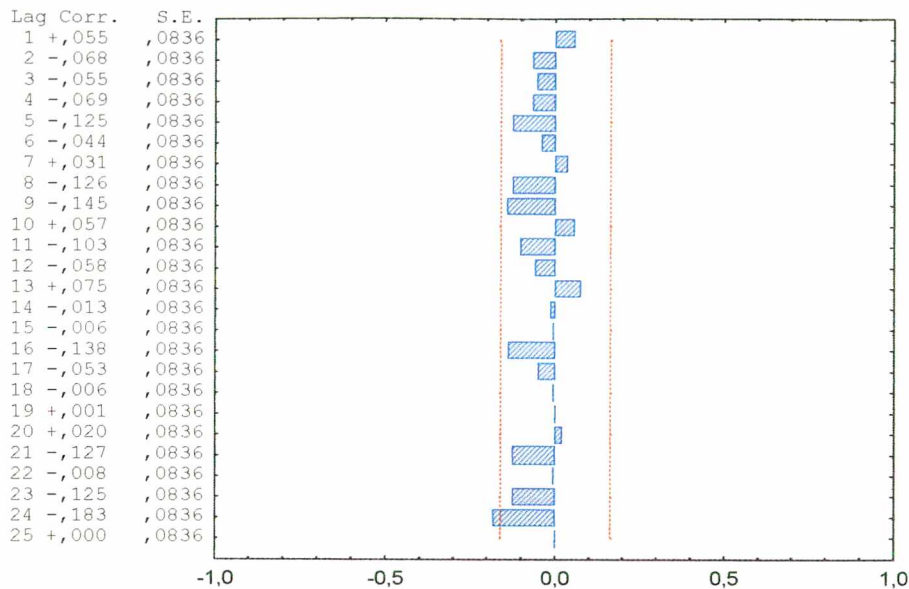


Figura 6.11 - Função de autocorrelação parcial (FACP) dos erros do modelo SARIMA (0,1,0)(2,1,0)₁₂.

A Fig. 6.10 e Fig. 6.11 demonstram que, com a aplicação do modelo, a autocorrelação nas lags cruciais foi eliminada e não há nenhum padrão discriminante em outras lags de ordem superior.

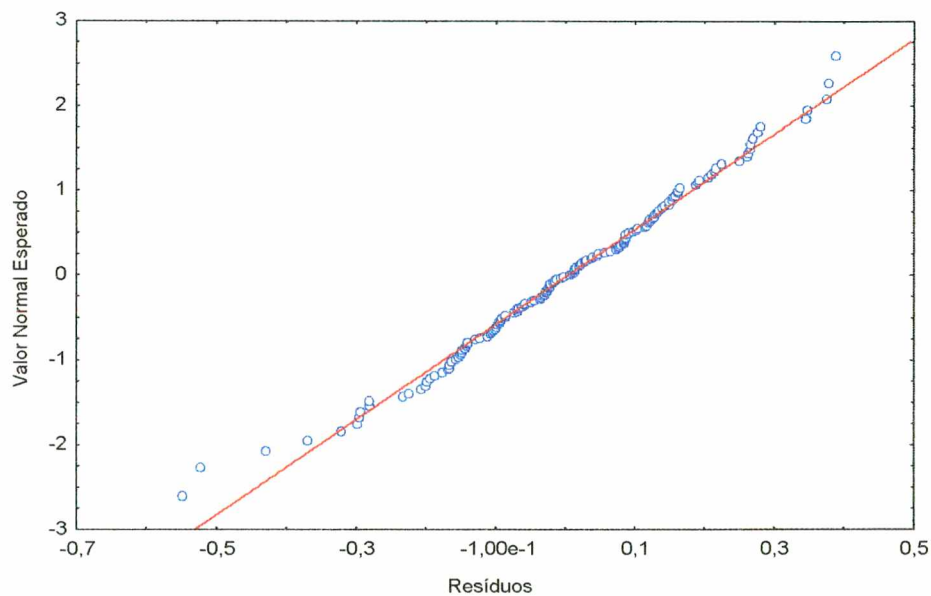


Figura 6.12 - Gráfico normal dos resíduos - modelo SARIMA (0,1,0)(2,1,0)₁₂.

Confirmando o que foi mostrado em BOX & JENKINS (1976), a Fig. 6.12 mostra que como o modelo fixado é adequado, então os resíduos estimados \hat{a}_t tendem

para um "ruído branco" a_t com média nula e variância σ_e^2 , com o crescimento de " n " (" n " sendo o número de termos da série diferenciada, isto é, $n = N - d$).

6.4 AVALIAÇÃO DA RNA

Como não foi encontrada uma metodologia padrão para avaliação da qualidade das previsões, optou-se, neste trabalho, pelo método utilizado por MAKRIDAKIS et al. (1998), ou seja, utilizou-se a maioria das medidas padrões de erro. Isso permite que outros autores possam comparar os resultados obtidos neste trabalho com os seus.

As previsões obtidas pelas RNAs foram comparadas com as provenientes do modelo SARIMA, através dos seguintes indicadores: Erro Médio (ME), Erro Absoluto Médio (MAE), Erro Quadrático Médio (MSE), Erro Quadrático Médio Normal Absoluto (NRMSE), Erro Percentual Absoluto Médio (MAPE) e o coeficiente de desigualdade (U_2) de Theil. Todos descritos no Item 2.1.4.

6.5 RESULTADOS OBTIDOS

Os resultados obtidos, podem ser visualizados na tabela 6.2.

Tabela 6.2 - Comparação dos resultados obtidos.

Medida de Exatidão	Método Trivial	RNA	SARIMA
ME	0,141	-0,135	2,626
MAE	1,482	0,981	2,790
MSE	3,269	1,718	13,469
NRMSE	1,808	1,310	1,609
MAPE (%)	7,412	4,836	13,350
U_2	1,000	0,675	1,941

Como pode-se observar na Tabela 6.2, em todas as medidas, o erro cometido pela RNA foi menor. A medida MAPE, mais utilizada na literatura (ver Item 2.1.4), obtido pela aplicação das RNAs, ficou abaixo de 5%, enquanto que para o modelo SARIMA, o erro ficou próximo de 13,5%. A Estatística U_2 de Theil, sugerida por ARMSTRONG & COLLOPY (1992) para seleção do método de previsão mais preciso, ficou próxima de 0,675, comprovando a superioridade da RNA.

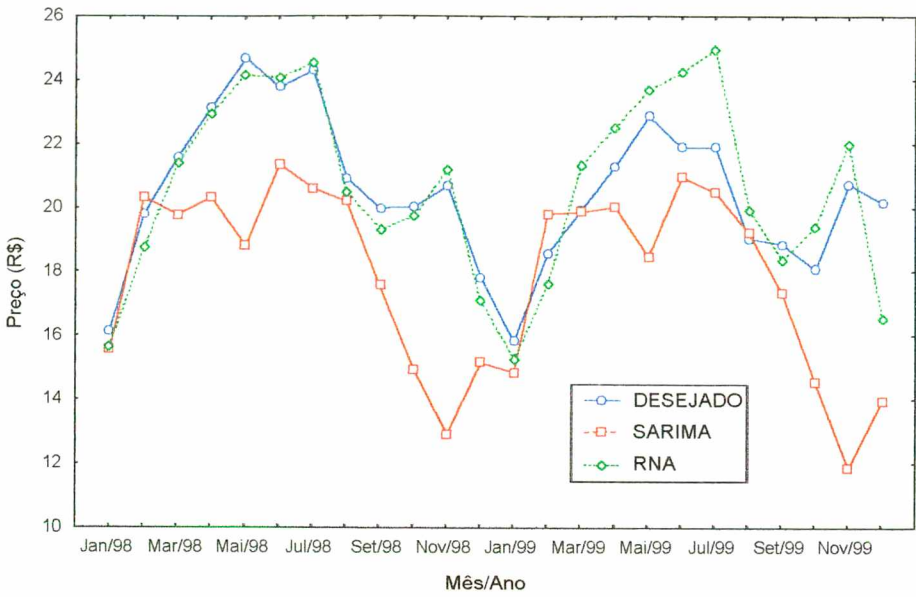


Figura 6.13 - Comparação dos resultados obtidos.

O fato da RNA ter conseguido resultados tão expressivos confirma sua capacidade de realizar prognósticos acertados. Os resultados obtidos pelas duas formas de previsão, ilustrados na Fig. 6.13, mostram que para este problema, em particular, tanto para previsões de curto, como de médio ou longo prazo, a RNA obteve melhores resultados.

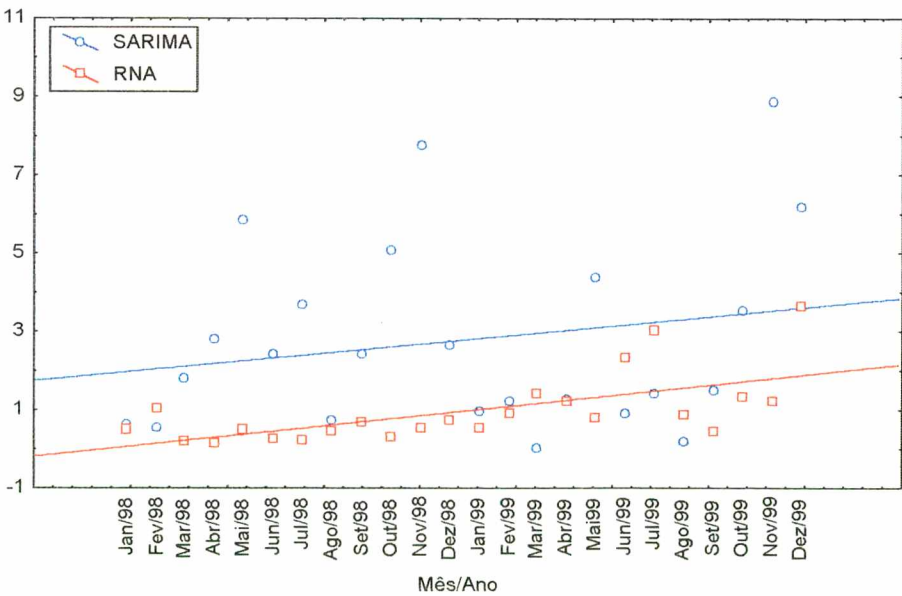


Figura 6.14 - Gráfico de dispersão dos erros obtidos, após a aplicação dos modelos.

A Fig. 6.14 mostra o gráfico de dispersão dos erros obtidos pelos dois modelos. Como pode-se notar, os erros cometidos pelo modelo SARIMA estão bem mais dispersos do que os erros cometidos pela RNA.

A Tabela 6.3 resume a estatística descritiva dos erros dos dois modelos. A média de erros da RNA é de 0,981, enquanto a do modelo SARIMA é de 2,790. Em termos de probabilidade, pode-se dizer que a probabilidade da RNA errar é de 0,260, enquanto a do modelo SARIMA é de 0,740.

Tabela 6.3 - Medidas descritivas dos erros cometidos pelos dois modelos no teste de previsão (1988 e 1999).

	SARIMA	RNA
Média	2,790	0,981
Mínimo - Máximo	0,053 - 8,871	0,168 - 3,656
Variância	5,934	0,789
Erro Padrão	0,464	0,169
Probabilidade de Errar	0,740	0,260

A Tabela 6.3 deixa claro que, na amostra observada, em média os erros obtidos pela aplicação da RNA foram menores que os obtidos pelo modelo ARIMA.

CAPÍTULO 7 - ANÁLISE E DISCUSSÕES

Discutir a habilidade de representação, o processo de construção do modelo e a aplicabilidade das RNAs em comparação com o método de Box & Jenkins;

O estudo de caso feito neste trabalho mostra que as RNAs podem, realmente, fornecer boas previsões de séries temporais. Porém, o desempenho obtido pelas RNAs, principalmente de redes MLP, depende da topologia destas redes (tamanho, estrutura e conexões), assim como dos parâmetros do algoritmo de treinamento (taxa de aprendizado, termo de momento e outros). A determinação da rede (o que inclui a arquitetura e os parâmetros de aprendizado) afeta a velocidade e a eficiência do processo de aprendizado, a tolerância a falhas e ruídos e a capacidade de generalização. Para entender melhor como e por que as RNAs podem ser usadas como modelos para previsão de séries de tempo, serão discutidos abaixo vários fatores como: habilidade de representação, o processo de construção do modelo e a aplicabilidade das RNAs em comparação com o método de Box & Jenkins.

7.1 REPRESENTAÇÃO

Como visto na seção 3.1, os modelos de Box & Jenkins são uma família de modelos lineares Auto-Regressivos e de Médias Móveis. Para a série de preços de ovos, o modelo de Box & Jenkins identificado tem a seguinte forma:

$$(1 - B^{12})(1 - B)y_t = (1 - \Phi_1 B^{12} - \Phi_2 B^{24})y_t \quad (64)$$

Rescrevendo o modelo, tem-se:

$$(1 - B^1 - B^{12} + B^{13})y_t = (1 - \Phi_1 B^{12} - \Phi_2 B^{24})y_t \quad (65)$$

ou ainda,

$$y_t = y_{t-12} + (y_{t-1} - y_{t-13}) + (y_{t-1} - \Phi_{1,12}y_{t-12} - \Phi_{2,24}y_{t-24}) \quad (66)$$

Ou seja, a previsão do preço para o período t é a soma do: 1) preço do mesmo mês no ano anterior; 2) uma componente tendência determinado pela diferença do valor

do mês prévio e o valor do mês prévio no ano anterior; 3) e o efeito aleatório dos valores do período $t-1$, $t-12$ e $t-24$ da série.

Se a série temporal é determinada por um modelo linear como o descrito acima, a metodologia de Box & Jenkins pode prever bem, contanto que o padrão não mude. Porém, se a série é determinada por um processo não-linear, por exemplo, a série logística gerada por $y(t+1) = ay(t)(1-y(t))$, é provável que a metodologia de Box & Jenkins falhe, uma vez que não existe no modelo termos de ordem maior. Por outro lado, uma RNA com uma única camada intermediária pode capturar a não-linearidade da série logística.

Uma RNA direta pode ser considerada como um modelo não-linear geral. Em efeito, esta é uma função complexa que consiste em um conjunto de funções de ativação $f \in C$ onde C é um conjunto de funções contínuas diferenciáveis, e um conjunto de parâmetros W chamado pesos. De acordo com TANG & FISHWICK (1991), a saída de uma RNA direta pode ser escrita como:

$$y = f\left(\sum_j w_{jk} f_j\left(\sum_m w_{mj} f_m\left(\cdots f_l\left(\sum_i w_{il} x_i\right)\right)\right)\right) \quad (67)$$

onde x_i é o $i^{\text{ésimo}}$ elemento do vetor de entrada x .

Foi provado por CYBENKO (1988), apud BRAGA et al (2000), p. 49, que - uma rede com uma camada intermediária pode implementar qualquer função contínua. A utilização de duas camadas intermediárias permite a aproximação de qualquer função ((CYBENKO, 1989) apud (BRAGA et al., 2000), p. 49). Estes resultados teóricos fazem das RNA diretas boas candidatas para o mapeamento de qualquer função complexa. Segundo TANG & FISHWICK (1991), as RNAs têm potencial para representar qualquer mapeamento complexo, não-linear, que pode governar mudanças em uma série temporal.

De acordo com DE AZEVEDO (2000), sem uma camada intermediária, a RNA se torna uma função da combinação linear das variáveis de entrada.

$$y = f\left(\sum_i w_{il} x_i\right) \quad (68)$$

Neste caso, x_i é definido como x_{t-i} , assim a (Equação 68) é semelhante ao modelo de Box & Jenkins, exceto que 1) contém uma função de ativação não-linear $f(x)$, e 2) não contém nenhum ruído branco. Essas diferenças, porém, podem ser desprezíveis devido ao fato da função de ativação poder ser mudada para uma função linear. As funções semilineares tornam-se linear quando $k \rightarrow \infty$ (ver Item 5.1.2.2), e novas unidades de entrada, que levam em conta erros aleatórios, podem ser criadas. Se for considerado uma RNA direta (com camadas intermediárias) que tenha conexões diretas das unidades de entrada para as unidades de saída, tem-se um modelo que combina: um modelo essencialmente linear (conexões diretas) e um modelo não-linear (através da camada intermediária). Consequentemente, pode-se concluir que modelos de RNAs são super conjuntos de modelos Box & Jenkins. Sendo assim, uma RNA têm habilidade para lidar com séries temporais mais complicadas.

7.2 ROBUSTEZ DO MODELO DE RNA

Com configurações diferentes dos parâmetros de treinamento, os modelos de RNAs produzem boas previsões. Embora aja muitos fatores que afetem o desempenho de modelos de RNAs (ver Item 5.1.1), parece existir uma permuta entre esses fatores. Por exemplo, os erros de previsão são grandes para muitas séries quando os parâmetros de treinamento são pequenos. Porém, estendendo a duração do treinamento, freqüentemente os erros de previsão serão reduzidos. A estrutura da RNA (número de unidades intermediárias) tem efeito no desempenho da previsão, porém, até mesmo com estruturas fixas - como usadas por muitos pesquisadores - foi mostrado que as RNAs podem competir com os modelos convencionais. Assim sendo, as RNAs são bastante robustas.

A robustez das RNAs também é refletida pelo fato destas serem modelos essencialmente livres de suposições, embora algumas suposições sobre o conjunto de treinamento e algumas inferências estatísticas possam ser realizadas. Esta propriedade torna as RNAs aplicáveis para um grande número de problemas de reconhecimento de padrões. O modelo de Box & Jenkins, como outros modelos com base estatística, estão sujeitos as suposições sobre a série de dados (ex. a distribuição de erros deve seguir uma distribuição normal).

7.3 GENERALIZAÇÃO

O número de neurônios da camada de entrada e de saída, é definido pelo problema. Já, o número de neurônios da camada intermediária é, em geral, definido empiricamente. Neurônios demais na camada intermediária pode levar a rede a memorizar os padrões de treinamento, em vez de extrair deles as características gerais que permitirão a generalização. Por outro lado, um número de neurônios muito pequeno, pode forçar a rede a gastar tempo em excesso tentando encontrar uma representação ótima, quando não é possível, sobrecarregando os neurônios.

A metodologia de Box & Jenkins evita o problema de *overfitting* pelo diagnóstico no processo de validação do modelo. O número de parâmetros nos modelos de Box & Jenkins é controlado de maneira que apenas aqueles parâmetros que contribuem para o ajuste dos dados com um certo grau de significância estatística são retidos. No caso das RNAs, não há nenhum procedimento estabelecido para prevenir o problema de *overfitting*, embora várias técnicas tenham sido sugeridas na literatura (BRAGA et al., 2000).

Uma delas é estimar o erro de generalização durante o processo de treinamento. Divide-se o conjunto de dados em três conjuntos (treinamento, teste e validação). O treinamento deve ser interrompido quando o erro do conjunto de validação começar a subir, isto é, quando a rede começar a incorporar o ruído presente nos dados. Segundo BRAGA et al. (2000), embora eficiente em algumas situações, o uso dessa alternativa é limitada quando o conjunto de dados não é muito grande, já que os dados do conjunto de validação não podem ser utilizados para treinamento.

Uma outra solução seria a adoção de técnicas de *pruning*, que envolvem a eliminação de pesos e nodos irrelevantes para a função executada pela rede. Um dos métodos de *pruning*, de acordo com BRAGA et al. (2000), é feito pela retirada de elementos da rede e a subsequente verificação do erro de saída. Caso a retirada do elemento não cause grande variação no erro de saída, este pode ser retirado sem que a rede perca sua capacidade de modelar os dados.

Como desvantagem do método, BRAGA et al. (2000) cita o fato de que a retirada dos elementos não considera a correlação entre eles. Um par de neurônios que

anula suas saídas em uma camada intermediária da rede não tem nenhum efeito na saída da rede, mas individualmente podem ter um grande efeito quando a remoção de um deles é analisada.

7.4 APRENDIZADO EM RNAs

O poder de representação das RNAs não pode ser utilizado a menos que algoritmos de aprendizagem eficientes existam. O algoritmo de retropropagação do erro (*Backpropagation*) utilizado no treinamento de RNAs diretas MLP é essencialmente um algoritmo baseado na descida do gradiente. Os dois principais problemas do algoritmo Backpropagation são: (1) convergência lenta e (2) mínimos locais. O primeiro problema foi investigado por muitos pesquisadores e muito progresso foi feito (BRAGA et al., 2000). Menor progresso foi alcançado no estudo do segundo problema. Porém, este problema, segundo TANG & FISHWICK (1991), é raro para a maioria das aplicações e pode ser aliviado alterando-se a arquitetura da rede.

A maior parte do tempo dedicado à construção de uma RNA é gasta no processo de treinamento. Muitas vezes, é preciso fazer alguns experimentos para determinar uma boa arquitetura de RNA, ou seja, um modelo que seja capaz de abstrair os dados com precisão sem incluir ruído. Porém, como mostrado no Item 5.2.2 os modelos de RNA são bastante robustos.

A escolha de uma RNA é relativamente fácil se comparada à construção de um modelo SARIMA. O processo de identificação e validação do modelo SARIMA necessita de conhecimento *à priori* do especialista. Além disso, não é fácil compreender o mecanismo estatístico usado no procedimento de identificação do modelo. No sentido de que menos conhecimento estatístico e interação com o usuário são requeridos, os modelos de RNAs são relativamente mais fáceis de se usar do que a metodologia Box & Jenkins.

Porém, em comparação à metodologia Box & Jenkins, a aproximação por RNAs não está completamente madura. Ao contrário da metodologia Box & Jenkins, falta um procedimento sistemático para determinação da RNA (o que inclui a arquitetura e os parâmetros de aprendizado). Frequentemente, métodos *ad hoc* ou até mesmo estruturas de RNAs e procedimentos de treinamento arbitrários são usados. Devido à robustez dos

modelos, ainda assim, um bom desempenho pode ser obtido, porém, a falta de uma base teórica e de procedimentos sistemáticos na construção dos modelos dificulta a aplicabilidade destes. Visando resolver este problema de otimização de arquiteturas, várias técnicas para automação do projeto de RNAs para classes particulares de problemas vêm sendo pesquisadas. BRAGA et al. (2000) agrupa a essas técnicas em três abordagens:

- otimização genética (abordagem evolucionária)¹;
- técnicas de *pruning*, como discutido anteriormente e,
- redes construtivas, onde utiliza-se uma rede inicialmente pequena, na qual novas unidades são inseridas durante o treinamento.

Uma abordagem mais detalhada sobre as três abordagens, pode ser encontrada em BRAGA et al. (2000).

Outra desvantagem das RNAs é que a rede treinada não fornece muita informação sobre a estrutura subjacente (função) de uma série temporal. Em outras palavras, se comparadas à metodologia Box & Jenkins que fornece equações lineares simples, nas RNAs faltam a habilidade para elucidar ou descrever de forma concisa a série temporal aprendida. Entretanto, esta negligência poderia ser superada. Recentemente, muitos pesquisadores estão estudando a extração de regras (ex. regras de produção) de uma RNA treinada (ver BRASIL, 1999).

¹ Ver ROISENBERG (1998).

CAPÍTULO 8 - CONCLUSÕES FINAIS

*Interpretar e analisar os dados obtidos visando atender aos objetivos da pesquisa;
Sintetizar os resultados obtidos.*

Neste estudo apresenta-se um modelo de RNA para predição do preço do ovo pago ao produtor.

O objetivo principal era criar um modelo capaz de prever o preço do ovo para um período mínimo de 10,6 meses à frente (período de formação das aves ao pico da produção). Tendo o conhecimento do preço futuro, o avicultor pode estimar a rentabilidade do empreendimento futuro e assim, organizar e programar a produção e a comercialização, reduzindo os riscos.

O modelo proposto, uma RNA direta (RNA(13, 10, 1) TO_Delay Estática), apresentou um ótimo desempenho. Nos testes realizados pôde-se prever com precisão o preço do ovo para um período de 14 meses à frente.

Para avaliar o desempenho da rede utilizou-se várias medidas de erro: ME (-0,135), MAE (0,981), MSE (1,718), NRMSE (1,310), MAPE (4,836%) e U_2 (0,675), sendo que todas comprovaram o bom desempenho da RNA.

Além disso, fatores como: a capacidade representativa, o processo de identificação e a capacidade preditiva da RNA foram comparadas com a metodologia Box & Jenkins para modelos univariados.

Apesar da metodologia Box & Jenkins não ser comparável teoricamente às RNA (essa trata-se de um modelo linear), essa é muito utilizada para a previsão de séries econômicas, como o preço do ovo.

Essa série é considerada complexa, dado o caráter aleatório dos fenômenos climáticos e biológicos que afetam a produção, e o grande número de variáveis que afetam as oportunidades de comercialização e lucro do produtor.

Um fator importante é a facilidade na identificação de uma RNA quando comparada a identificação do modelo SARIMA. Em um modelo SARIMA, tenta-se obter os estimadores das funções teóricas através de uma realização do processo estocástico. As funções amostrais nem sempre são bem comportadas quanto as funções teóricas, tornando o processo de identificação complicado.

Na identificação da RNA o único conhecimento dos dados necessário é a sazonalidade da série. Com base nesta, fica determinado o número de neurônios da camada de entrada da rede. O número de neurônios da camada de saída depende do tipo de previsão desejada (*single* ou *multi-step*). O mais difícil talvez, seja a determinação do número de neurônios da camada intermediária. Na maioria das vezes este é determinado por heurística ou regras *ad-hoc*. Porém, já existem métodos que através da abordagem evolucionária, tentam determinar o número ótimo de neurônios dessa camada (BARRETO, 1996).

Outra facilidade das RNAs é que estas são, em essência, modelos não-paramétricos, ou seja, modelos que não especificam condições sobre os parâmetros da população da qual se extraiu a amostra. Um conjunto de exemplos é apresentado para a rede, a qual extrai automaticamente as características necessárias para representar a informação fornecida.

Já modelos paramétricos como o modelo SARIMA, especificam certas condições como a de que a série em estudo deve ser estacionária na média e variância. Além disso, a construção do modelo depende de um entendimento mais profundo do problema, e dependendo do número de variáveis consideradas, pode ter alta complexidade.

Porém, apesar da natureza indutiva das RNAs, como mostrado por ZHANG & HUTCHINSON (1994), BALLINE et al. (1997), FARAWAY & CHATFIELD (1998) e neste trabalho, o conhecimento à prior dos dados e do problema a ser tratado melhora significativamente o desempenho das RNAs, podendo até ser indispensável para se chegar a uma solução satisfatória.

A flexibilidade das RNA's permite o uso tanto do conhecimento intuitivo do especialista com informações geradas por modelos estatísticos formais.

A observação da FAC e FACP (procedimento usual na metodologia de Box & Jenkins), modelos estatísticos paramétricos, ajuda a determinar quantas e quais observações passadas são importantes para a previsão de valores futuros. Essa informação pode ser usada na definição da topologia da rede.

De todos estes itens, talvez o mais importante para um bom desempenho em previsão de séries temporais, como citado por ZHANG & HUTCHINSON (1994) é conhecer os dados e a técnica de modelagem.

Outro ponto importante é a configuração do algoritmo de treinamento. Uma configuração inadequada pode fazer com que a rede não aprenda. Isso pode explicar a diversidade de resultados encontrados na literatura.

As medidas escolhidas para avaliar o desempenho da RNA, quando o problema é de predição, também podem levar a resultados divergentes. Uma saída seria usar diversas medidas, como feito neste trabalho.

A capacidade de aprender através de exemplos e de generalizar a informação aprendida é, sem dúvida, o atrativo principal da solução de problemas através de RNAs.

Uma RNA é capaz de extrair informações não apresentadas de forma explícita através dos exemplos. Não obstante, são capazes de atuar como mapeadores universais de funções multivariáveis (BRAGA et al., 2000).

Enfim, este trabalho não tem a pretensão de estabelecer que as RNAs, mais especificamente o modelo *Perceptron Multi-Camadas* com aprendizado por retropropagação de erros, apresentam-se como o melhor método de realização de previsão de valores futuros de séries temporais, posto que o trabalho empírico não é abrangente e os modelos não são comparáveis teoricamente. Porém, em face dos resultados apresentados, pode-se dizer que as RNAs são uma ferramenta poderosa para a realização de previsão de séries temporais.

CAPÍTULO 9 - TRABALHOS FUTUROS

Sugere-se como proposta de trabalhos futuros o desenvolvimento de um sistema de simulação da taxa de produção de ovos em função do tempo. Um bom modelo seria o proposto por FIALHO & LEDUR (1997). Com o sistema poderia-se prever a produção futura de ovos que em conjunto com o modelo de predição de preço proposto neste trabalho, pode estimar a rentabilidade de uma unidade de produção de ovos, o custo e o tempo necessário para recuperar o capital investido.

Para tentar melhorar ainda mais a previsão obtida neste trabalho, poderia-se testar outros tipos de memória de curto-tempo, como uma memória TOS ou TIS exponencial, ou ainda, tornar essa memória dinâmica.

Outra sugestão para melhorar o desempenho tanto da RNA como do modelo SARIMA, seria testar a influência da ocorrência de planos econômicos na qualidade da previsão. Só no período da série em estudo (1985 - 1999), o sistema monetário passou por cinco reformas. Acredita-se que estas podem influenciar bastante na qualidade da previsão.

Outra lacuna a ser preenchida, essa na área de RNAs, é a determinação do número de neurônios da camada intermediária de uma rede. O que existe até o momento são regras *ad-hoc* que às vezes funcionam para certos tipos de rede. Porém, nenhuma solução plausível foi obtida até o momento para este problema.

Para finalizar, este trabalho é só um exemplo da aplicação de métodos de Inteligência Computacional na resolução de problemas reais. Muitos outros problemas considerados sem solução ou com soluções não plausíveis poderão ser resolvidos em um futuro próximo, com a aplicação desse novo paradigma.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABELÉM, A.J.G. **Redes Neurais Artificiais na Previsão de Séries Temporais**. Rio de Janeiro, 1994. Dissertação (Mestrado em Engenharia Elétrica) - Departamento de Engenharia Elétrica, Potifícia Universidade Católica do Rio de Janeiro.
- AHLBURG, D. A. Error measures and the choice of a forecast method. In: **A commentary on error measures**. International Journal of Forecasting, North-Holland, v. 08, p.99-111, 1992.
- ANDRADE, L.C.M.; DINIZ, H.; CARVALHO, A.C.P.L.F.; ANDRADE, M. G. **Previsão de séries temporais através de redes neurais artificiais e métodos estocásticos**. In: Simpósio Brasileiro de Redes Neurais (IV.: Dez/97: Goiânia). Anais... Goiânia, 1997, p. 13-17.
- ARMSTRONG, J.S.; COLLOPY, F. **Error measures for generalizing about forecasting methods: Empirical comparisons**. International Journal of Forecasting, North-Holland, v. 08, p.69-80, 1992.
- BALLINI, R.; FRANÇA, E.H.F.; SOARES, S.; ANDRADE, M.G. **Relação entre modelos auto-regressivos e a configuração da rede neural para previsões de séries temporais estacionárias**. In: Simpósio Brasileiro de Redes Neurais (IV.: Dez/97: Goiânia). Anais... Goiânia, 1997.
- BARRETO, J.M. **Inteligência Artificial: No limiar do século XXI**. Florianópolis: Duplic, 1ª edição, 1997.
- BOX, G.E.P.; JENKINS, G.M. **Time Series Analysis: Forecasting and control**, San Francisco: Holden-Day, 2ª edição, 1976.
- BRAGA, A. P.; CARVALHO, A.P.L.; LUDERMIR, T. B. **Redes neurais artificiais: teoria e aplicações**. Rio de Janeiro: Livros Técnicos e Científicos, 2000.

- BRAGA, M. J.; PAIVA, J.G.C; MACIEL, V.V. **O emprego de redes neurais na previsão de preços de café.** Trabalho apresentado ao VIII Simpósio Técnico de Produção de Ovos, São Paulo, 1999.
- BRAGA, M.J.; RESENDE, M. A. **Comparação entre o uso de redes neuronais e metodologia de Box & Jenkins na previsão de preços agrícolas: Um exercício empírico.** In: Congresso da sociedade Brasileira de Economia e Sociologia Rural, 34, 1996, Aracaju. Anais... Aracaju: SOBER, 1996. V. II, p. 1130-1148.
- BRASIL, L.M. **Proposta de arquitetura para sistema especialista híbrido e a correspondente metodologia de aquisição do conhecimento.** Florianópolis, 1999. Tese (Doutorado em Engenharia Elétrica) - Centro Tecnológico, Universidade Federal de Santa Catarina.
- CHAKRABORTY, K. et. al. Forecasting the behavior of multivariate time series using neural networks. **Neural Networks**, EUA, n. 5, p. 961-970, 1992.
- CORRÊA, W.R.; PORTUGAL, M.S. **Previsão de Séries de tempo na presença de mudança estrutural: Redes neurais artificiais e modelos estruturais.** Trabalho apresentado ao XVIII International Symposium on Forecasting, Edinburgh, Escócia, 1995.
- Curso de RNA. [on line]. Disponível: <http://www.furb.rct-sc.br/~malcon/SumárioCursoRN.htm> [capturado em 26 Mar. 01].
- DE AZEVEDO, F.M. **Algoritmos genéticos em redes neurais artificiais.** Trabalho apresentado ao IV Congresso Brasileiro de RN, São José dos Campos, 1999.
- DE AZEVEDO, F.M. Notas de aulas da disciplina Introdução as Redes Neurais Artificiais. Mestrado em Engenharia Biomédica - INE - UFSC, 2000.
- DE OLIVEIRA, C. M. **Um modelo de rede neural artificial para previsão automática da carga ativa nodal de curto prazo em sistemas de transmissão e distribuição de energia elétrica com supervisão e controle automatizados.** Florianópolis, 1999. Dissertação (Mestrado em Engenharia) - Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina.

- FARAWAY, J. CHATFIELD, C. **Time series forecasting with neural networks: a comparative study using the airline data.** Applied Statistics, São Paulo, v. 47, part 2, p.231-250, 1998.
- FAVA, V. L. **Análise de séries de tempo.** In: VASCONCELLOS, M.A.S.; ALVES, D. **Manual de Econometria.** São Paulo: Ed. Atlas, 2000a. P.199-205.
- FAVA, V. L. **Metodologia de Box-Jenkins para modelos univariados.** In: VASCONCELLOS, M.A.S.; ALVES, D. **Manual de Econometria.** São Paulo: Ed. Atlas, 2000b. P.205-231.
- FERNANDES, L.G.L. **Utilização de Redes Neurais na Análise e Previsão de Séries Temporais.** Porto Alegre, 1995. Dissertação (Mestrado em Ciências da Computação) - Curso de Pós-Graduação em Ciências da Computação, Universidade Federal do Rio Grande do Sul.
- FERNANDES, L.G.L.; PORTUGAL, M.S.; NAVAUX, P.O.A. **Previsão de Séries de Tempo: Redes Neurais Artificiais e Modelos Estruturais.** No prelo.
- FERNANDES, L.G.L.; PORTUGAL, M.S.; NAVAUX, P.O.A. **Um estudo de poder preditivo das redes neurais artificiais comparado a métodos econométricos tradicionais.** Anais do Primeiro Simpósio Brasileiro de Redes Neurais, Caxambu/MG, p. 139-144.
- FIALHO, F. B.; LEDUR, M. C. **Segmented polynomial model for estimation of egg production curves in laying hens.** British Poultry Science, v.38, n.1, p.66--73, 1997.
- Forecasting Principles Website.** [on line]. Disponível: <http://hops.wharton.upenn.edu/forecast> [capturado em 26 Mar. 01].
- FREEMAN, J.A.; SKAPURA, D.M. **Neural Networks - Algorithms, Applications, and Programming Techniques.** California: Addison-Wesley Publishing Company, 1992.
- HUNTON, P. **Analisando o mercado mundial do ovo.** Trabalho apresentado ao VIII Simpósio Técnico de Produção de Ovos, São Paulo, 1999.

- Introduction to ARIMA: Nonseasonal models.** [on line]. Disponível: www.geocities.com/Colosseum/5585/ARIM1.html [capturado em 26 Mar. 01].
- KADOWAKI, M.; SOARES, S.; ANDRADE, M. G. **Previsão de vazões mensais utilizando redes neurais multicamadas com algoritmo Backpropagation.** In: Simpósio Brasileiro de Redes Neurais (IV.: Dez/97: Goiânia). Anais... Goiânia, 1997, p. 32-35.
- KOVÁCS, Z.L. **Redes neurais artificiais – Fundamentos e aplicações.** São Paulo: Edição Acadêmica - USP, 1996.
- KREMER, S. C. **Spacio-temporal Connectionist Networks: A taxonomy and review.** [on line]. Disponível: <http://hebb.cis.uoguelph.ca/~skremer/Teaching/27642/dynamic2/review.html> [capturado em 26 Mar. 01].
- KROSE, B.J.A.; VAN DER SMAGT, P.P. **An Introduction to Neural Networks.** 7^a ed., 1995.
- LIMA Jr., J.A.; DE CARVALHO, A.C.P.L. **Utilização de Redes Neurais Artificiais para previsão de séries temporais.** In: Simpósio Brasileiro de Redes Neurais (IV.: Dez/97: Goiânia). Anais... Goiânia, 1997, p. 93 e 94.
- LOESCH, C.; SARI, S.T. **Redes neurais artificiais – Fundamentos e modelos.** Blumenau: Editora da FURB, 1996.
- MAKRIDAKIS, S.; ANDERSEN, A.; CARBONE, R.; FILDES, R.; HIBON, M.; LEWANDOWSKI, R.; NEWTON, J.; PARZEN, E. WINKLER, R. **The accuracy of extrapolation (time series) methods: results of a forecasting competition.** Journal of Forecasting, 1, 111-153, 1982.
- MAKRIDAKIS, S.; WHEELWRIGHT, S.C.; HYNDMAN, R.J. **Forecasting - Methods and applications.** Ed. 3. New York: John Wiley & Sons, Inc., 1998.
- MARTINS, S.S. **Cadeias produtivas do frango e do ovo: avanços tecnológicos e sua apropriação.** São Paulo, 1996. Tese (Doutorado em Economia) - Economia de Empresas, Escola de Administração de Empresas de São Paulo da Fundação Getúlio Vargas.

- MARTINS, S.S. **Mecanismos de coordenação nas cadeias produtivas do frango e do ovo.** Informações Econômicas, v.29, n. 6. São Paulo: IEA, 1999.
- MEHROTRA, K; MOHAN, C.K.; RANKA, S. **Elements of Artificial Neural Networks.** Cambridge: MIT, 1999.
- MILONE, G.; ANGELINI, F. **Estatística Aplicada.** São Paulo: Atlas, 1995.
- MOZER, M.C. Neural Net Architectures for Temporal Sequence Processing. In: WEIGEND, A.S.; GERSHENFELD, N. A. **Time series prediction - Forecasting the future and understanding the past.** Massachusetts: Perseus Books Publishing, 1994. P.243-264.
- MUELLER, A. **Uma aplicação de redes neurais artificiais na previsão do mercado acionário.** Florianópolis, 1996. Dissertação (Mestrado em Engenharia da Produção) - Centro Tecnológico, Universidade Federal de Santa Catarina.
- NORTH, M.O.; BELL, D.D. **Commercial Chicken Production Manual.** Ed. 4. New York: Chapman & Hall, 1990, 913p.
- PODDIG, T. **Short-Term forecasting of the USD/DM-exchange rate.** In: Neural Network in Capital Markets (1993: Londres). Anais... Londres: London Business School, 1993.
- PORTUGAL, M. S. **Neural networks versus time series methods: a forecasting exercise.** Revista Brasileira de Economia, Rio de Janeiro: FGV, v. 49, n. 4, p.611-629, 1995.
- REFENES, A. N. Currency exchange rate prediction and neural network design strategies. **Neural Computing & Applications Journal.** Londres, v. 1, n.1, p. 46-58, 1992.
- ROISENBERG, M. **Emergência da inteligência em agentes autônomos através de modelos inspirados na natureza.** Florianópolis, 1998. Tese (Doutorado em Engenharia Elétrica) - Centro Tecnológico, Universidade Federal de Santa Catarina.

- RUMELHART, D.E.; McCLELLAND, J. **Parallel Distributed Processing**, v. 1, MIT Press, 1986.
- RUMELHART, D.E.; WIDROW, B.; LEHR, M.A. **Artificial Intelligence**. Communications of the ACM, v. 37, n.3, 1994.
- SOUZA, R.C. **Modelos estruturais para previsão de séries temporais: abordagem clássica e bayesiana**. 17º Colóquio Brasileiro de Matemática. RJ: IMPA, 1989.
- SOUZA, R.C.; CAMARGO, M.E. **Análise e previsão de séries temporais: os modelos ARIMA**. Ijuí: SEDIGRAF, 1996.
- STERN, H.S. **Neural networks in applied statistics**. Technometrics, v. 38, n. 3, p.205-214, 1996.
- TAFNER, M. A; XEREZ, M.; RODRIGUES, I.W. **Redes Neurais Artificiais**. Blumenau: EKO, 1996.
- TAMARI, S.; RUIZ-SUDREZ, J.C.; WÖSTEN, J.H.M. **Testing an artificial neural network for predicting soil hydraulic conductivity**. In: Sixth International Conference on Computers in Agriculture (1996: Cancun). Anais. Cancun: American Society of Agricultural Engineers, 1996. P. 913-919.
- TANG, Z.; FISHWICK, P.A. **Feed-forward neural nets as models for time series forecasting**. Computer and Information Sciences, Florida, 1991.
- TAYLOR, S. J. Comparing forecasts in finance. In: **A commentary on error measures**. International Journal of Forecasting, North-Holland, v. 08, p.99-111, 1992.
- VALENÇA, M.J.S. Um estudo comparativo dos modelos Box & Jenkins, **Redes Neurais e Regressão Múltipla na previsão de vazões médias mensais**. In: Simpósio Brasileiro de Redes Neurais (IV.: Dez/97: Goiânia). Anais... Goiânia, 1997, p. 55-57.

- WAN, E.A. Time Series Prediction by Using a Connectionist Network with Internal Delay Lines. In: WEIGEND, A.S.; GERSHENFLD, N. A. **Time series prediction - Forecasting the future and understanding the past**. Massachusetts: Perseus Books Publishing, 1994. P.195-217.
- WEIGEND, A.S.; GERSHENFLD, N. A. The future of time series: learning and understanding In: WEIGEND, A.S.; GERSHENFLD, N. A. **Time series prediction - Forecasting the future and understanding the past**. Massachusetts: Perseus Books Publishing, 1994. P.1-70.
- YANG, C.C; LACROIX, R. PRASHER, S.O. **The use of back-propagation neural networks for the simulation and analyses of time-series data in subsurface drainage systems**. In: Sixth International Conference on Computers in Agriculture (1996: Cancun). Anais... Cancun: American Society of Agricultural Engineers, 1996. P. 940-949.
- ZHANG, G. P.; PATUWO, B. E.; HU, M. Y. **Nonlinear time series forecasting with artificial neural networks**. [on line]. Disponível: <http://fy.chalmers.se/~pin/exmain/node2.html> [capturado em 26 Mar. 01].
- ZHANG, X. & HUTCHINSON, J. Simple Architectures on Fast Machines: Practical Issues in Nonlinear Time Series Prediction. In: WEIGEND, A.S.; GERSHENFLD, N. A. **Time series prediction - Forecasting the future and understanding the past**. Massachusetts: Perseus Books Publishing, 1994. P.219-241.

ANEXO A

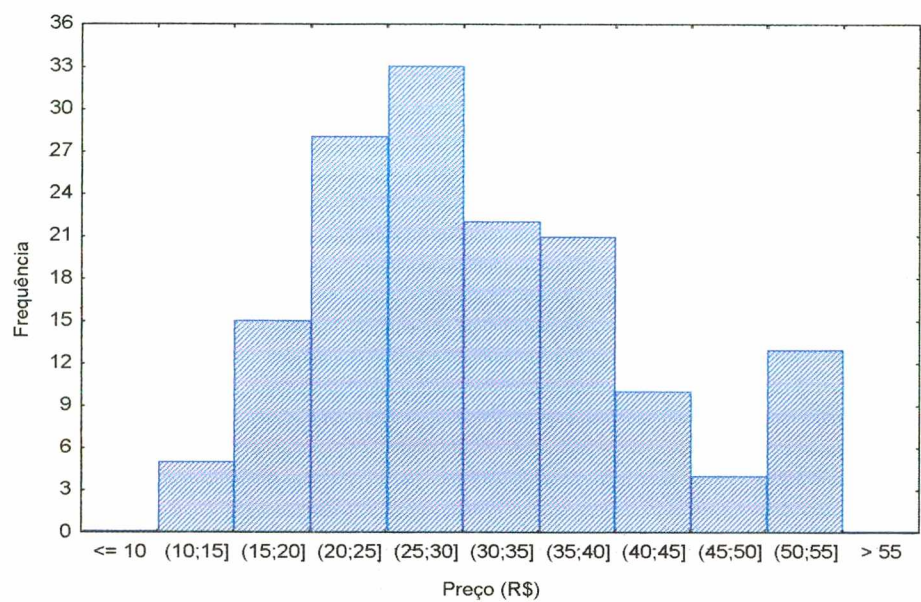


Figura A.1 - Histograma da variável em estudo (Preço do ovo branco (tipo grande) pagos ao produtor de São Paulo): Jan/85 a Dez/99.

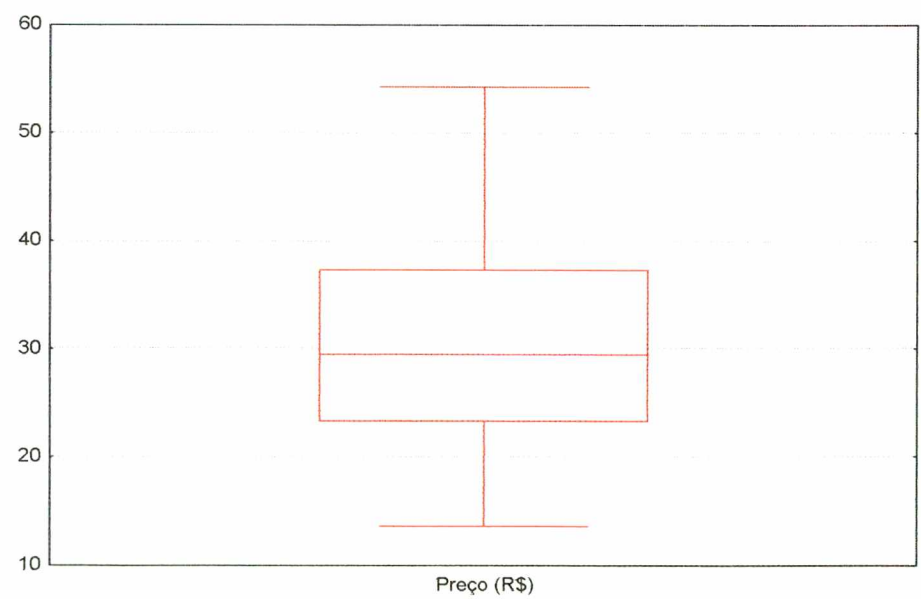


Figura A.2 - Representação da distribuição do preço do ovo em diagramas de caixas.

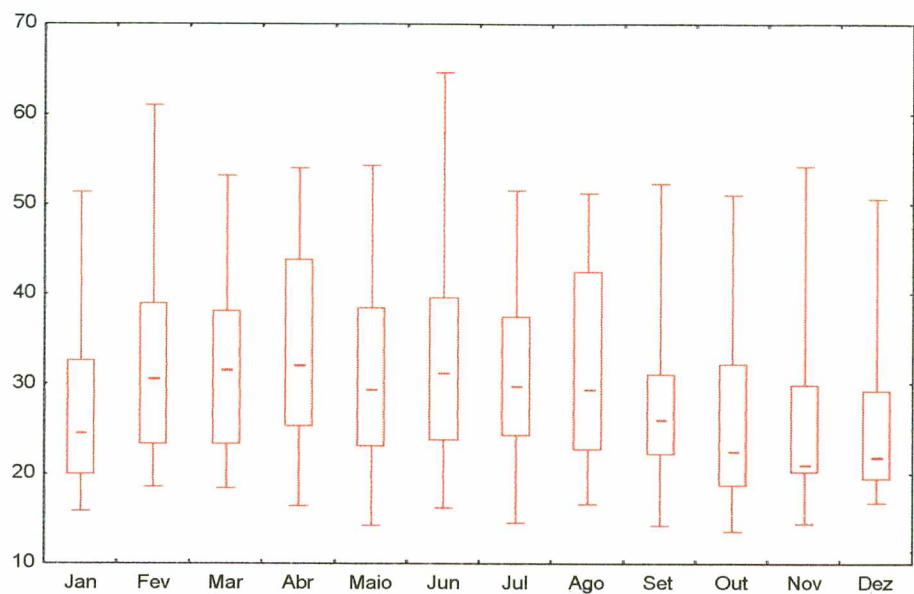


Figura A.3 - Representação da distribuição mensal do preço do ovo em diagramas de caixas.

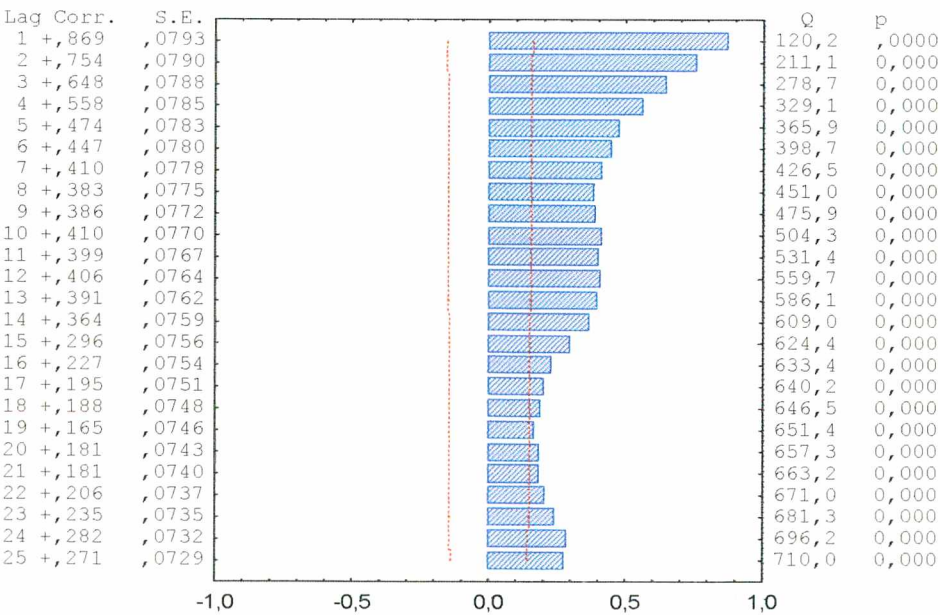


Figura A.4 - Função de Autocorrelação (FAC) da série transformada.

A FAC da série transformada (Fig. A.4) mostra uma queda não linear típica de uma série não-estacionária. As correlações decrescem lentamente para zero. Nota-se ainda, que trata-se de uma série sazonal (os valores de r_{12} e r_{24} são maiores que suas autocorrelações adjacentes. Segundo MAKRIDAKIS et al. (1998), isso sugere que seja aplicada uma diferenciação sazonal à série.

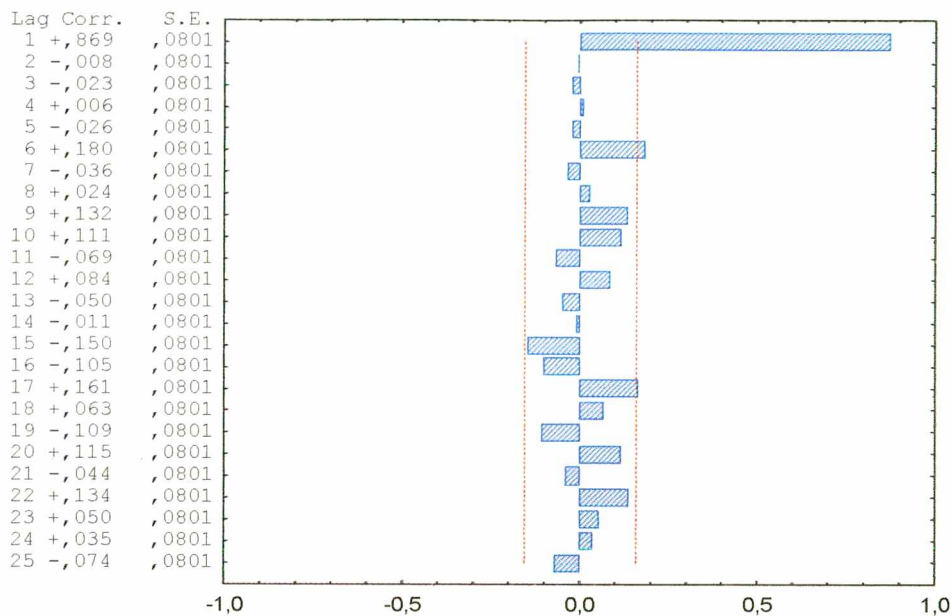


Figura A.5 - Função de Autocorrelação Parcial (FACP) da série transformada.

A FACP da série (Fig. A.5) mostra um pico significativo no *Lag* 1, significando que todas as auto correlações de ordem maior são explicadas efetivamente pela autocorrelação da *Lag* 1.

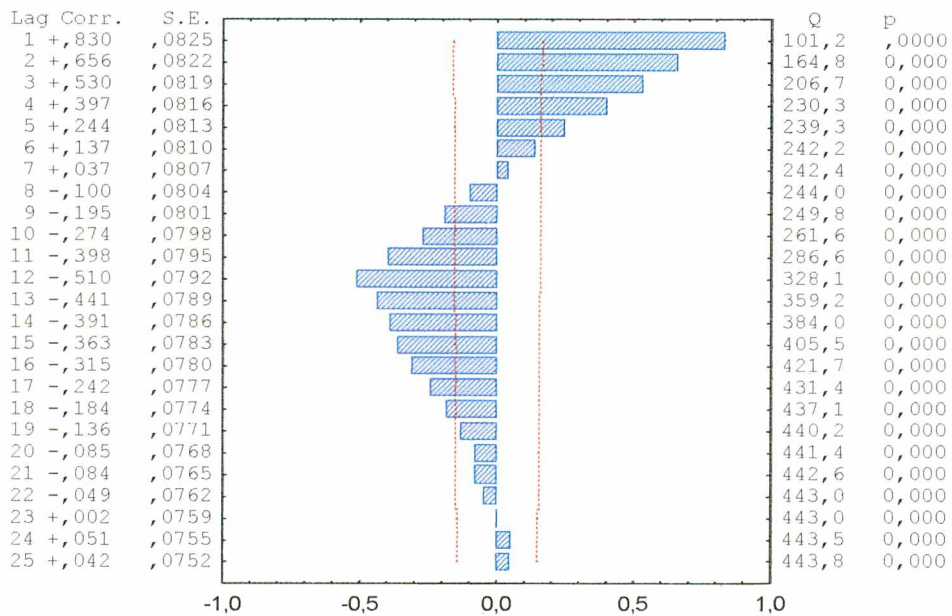


Figura A.6 - Função de Autocorrelação (FAC) da primeira diferença sazonal da série transformada.

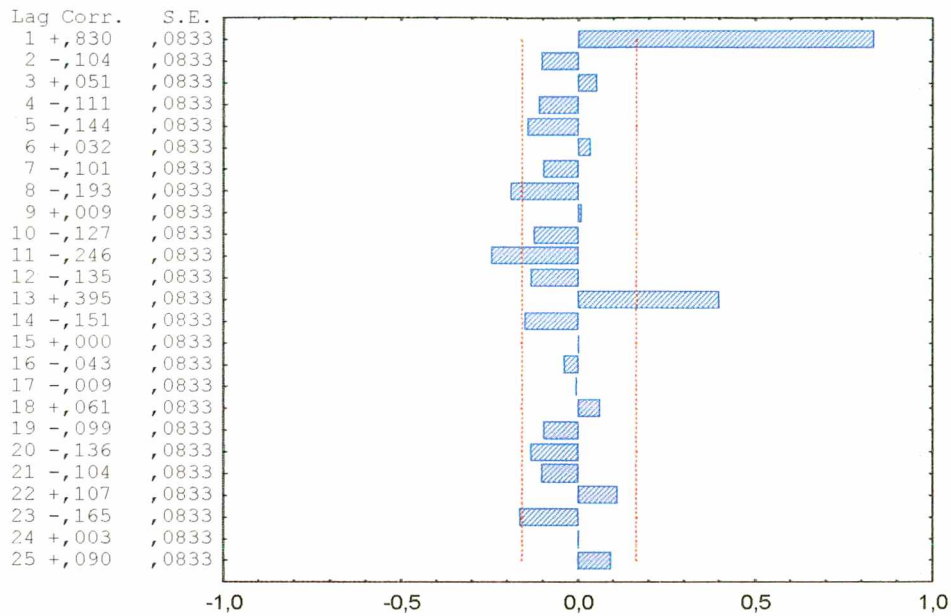


Figura A.7 - Função de Autocorrelação Parcial (FACP) da primeira diferença sazonal da série transformada.

As Figuras A.6 e A.7 mostram que a série ainda não está estacionária. As autocorrelações nas lags 1 e doze continuam salientes. A não-estacionariedade na média resultante pode ser removida com uma diferenciação primeira. A FAC e a FACP da série resultante é mostrada nas Figuras A.8 e A.9.

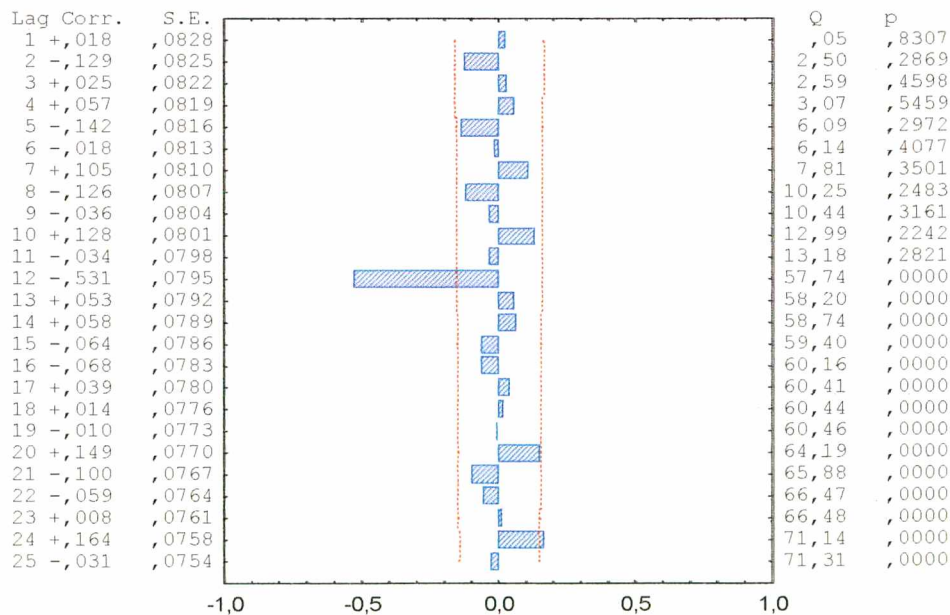


Figura A.8 - Função de Autocorrelação (FAC) da série após duas diferenciações.

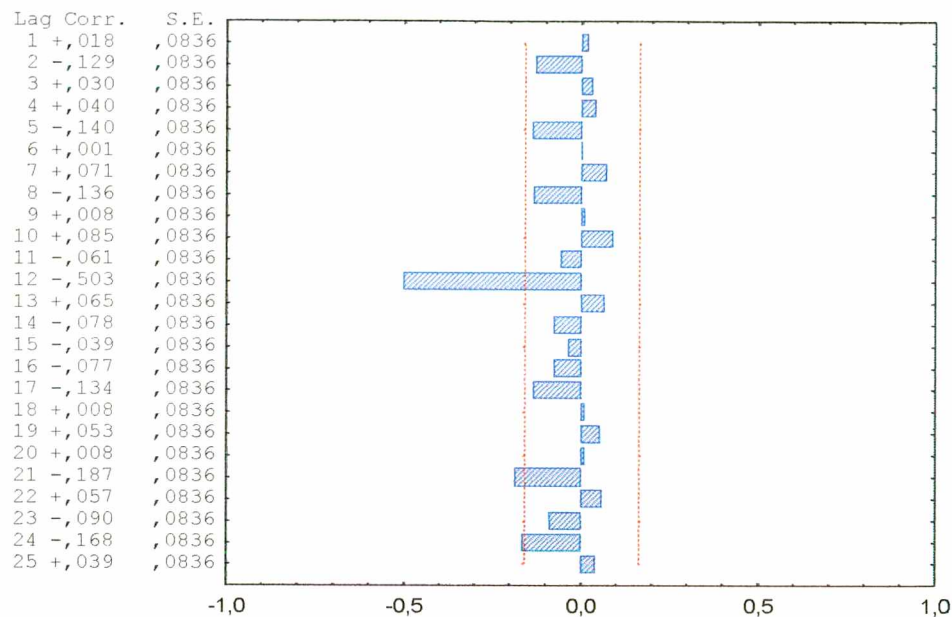


Figura A.9 - Função de Autocorrelação Parcial (FACP) da série após duas diferenciações.

Note que a série parece aproximadamente estacionária sem tendência longa: ela exibe uma tendência definitiva para retornar à média. O desvio padrão dos dados reduziu dramaticamente de 0,361 para 0,211 (Tabela A.1).

Na Fig. A.8 e A.9, observa-se claramente a saliência das autocorrelações multiplas de doze (*Lags* 12 e 24), o que indica que o modelo verdadeiro contém dois componentes AR sazonais - $AR(2)_{12}$.

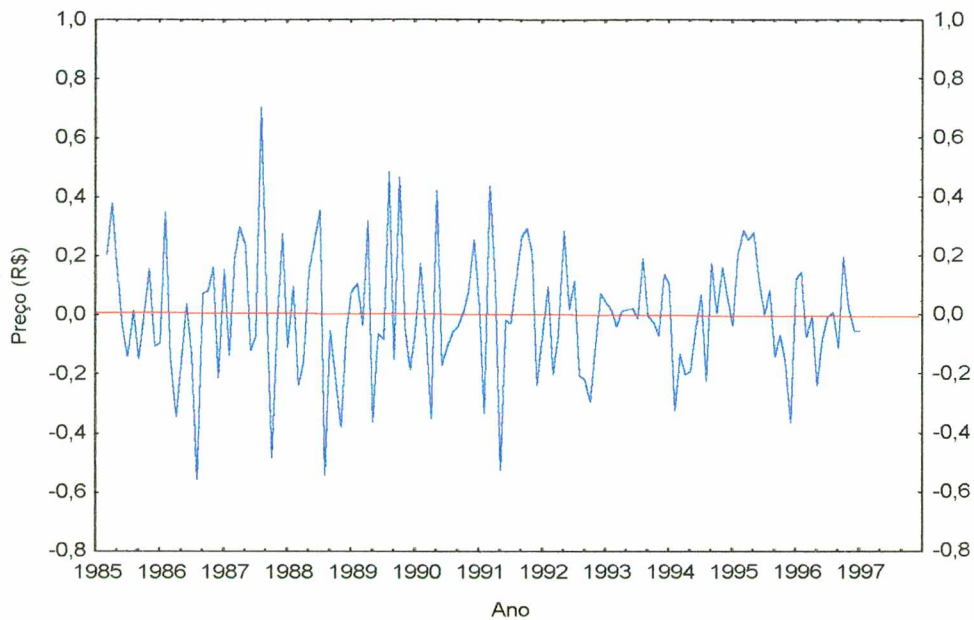


Figura A.10 - Resíduos da série diferenciada.

A Fig. A.10 apresenta os resíduos após as duas diferenciações. Como pode-se notar, os valores estão bem próximos de zero.

Tabela A.1 - Medidas descritivas da variável em estudo.

	Série			
	Original	Transformada	1ª Diferença Sazonal	2ª Diferença
Numero de elementos	153	153	141	151
Média	30,928	3,389	-0,050	0,000
Mínimo	13,618	2,611	-0,881	-0,557
Máximo	54,287	4,168	0,710	0,703
Desvio Padrão	10,258	0,346	0,361	0,211

ANEXO B

Tabela B.1 - Preço real do ovo branco (tipo grande) pagos ao produtor de São Paulo: Jan/85 a Dez/99.

Mês/Ano	Preço (R\$)	Valor ajustado conforme calendário	Transformação logaritma	Valor normalizado
jan/85	42,72	41,94	3,74	0,445
fev/85	43,30	47,07	3,85	0,593
mar/85	38,86	38,15	3,64	0,323
abr/85	32,98	33,46	3,51	0,155
mai/85	33,84	33,23	3,50	0,146
jun/85	39,01	39,58	3,68	0,370
jul/85	38,49	37,79	3,63	0,311
ago/85	44,47	43,66	3,78	0,497
set/85	44,18	44,82	3,80	0,530
out/85	38,09	37,40	3,62	0,298
nov/85	43,57	44,21	3,79	0,513
dez/85	46,22	45,38	3,82	0,546
jan/86	33,20	32,60	3,48	0,121
fev/86	41,29	44,88	3,80	0,532
mar/86	54,08	53,10	3,97	0,748
abr/86	53,18	53,96	3,99	0,769
mai/86	52,52	51,57	3,94	0,710
jun/86	52,42	53,18	3,97	0,750
jul/86	52,42	51,47	3,94	0,708
ago/86	52,18	51,23	3,94	0,702
set/86	51,51	52,26	3,96	0,728
out/86	51,93	50,99	3,93	0,696
nov/86	53,40	54,18	3,99	0,774
dez/86	51,49	50,56	3,92	0,685
jan/87	52,31	51,36	3,94	0,705
fev/87	56,10	60,98	4,11	0,926
mar/87	52,07	51,13	3,93	0,699
abr/87	43,89	44,53	3,80	0,522
mai/87	45,00	44,18	3,79	0,512
jun/87	39,06	39,63	3,68	0,372
jul/87	22,38	21,97	3,09	-0,385
ago/87	23,95	23,52	3,16	-0,298
set/87	25,70	26,07	3,26	-0,166
out/87	30,41	29,86	3,40	0,008
nov/87	25,26	25,63	3,24	-0,188
dez/87	28,35	27,84	3,33	-0,082
jan/88	25,06	24,61	3,20	-0,240
fev/88	33,64	35,31	3,56	0,224
mar/88	40,58	39,84	3,68	0,379
abr/88	43,13	43,76	3,78	0,499
mai/88	39,20	38,49	3,65	0,335
jun/88	31,72	32,18	3,47	0,105
jul/88	36,71	36,04	3,58	0,250
mar/92	32,20	31,62	3,45	0,082
abr/92	24,82	25,18	3,23	-0,210

Mês/Ano	Preço (R\$)	Valor ajustado conforme calendário	Transformação logaritma	Valor normalizado
ago/88	43,27	42,48	3,75	0,462
set/88	28,63	29,05	3,37	-0,027
out/88	32,85	32,25	3,47	0,108
nov/88	35,87	36,39	3,59	0,263
dez/88	36,01	35,36	3,57	0,226
jan/89	35,02	34,38	3,54	0,190
fev/89	35,77	38,88	3,66	0,348
mar/89	38,28	37,59	3,63	0,304
abr/89	47,05	47,74	3,87	0,611
mai/89	55,29	54,29	3,99	0,776
jun/89	63,68	64,61	4,17	1,000
jul/89	42,85	42,07	3,74	0,449
ago/89	47,87	47,00	3,85	0,591
set/89	25,26	25,63	3,24	-0,188
out/89	19,79	19,43	2,97	-0,543
nov/89	20,55	20,85	3,04	-0,453
dez/89	22,34	21,93	3,09	-0,388
jan/90	24,17	23,73	3,17	-0,287
fev/90	23,80	25,87	3,25	-0,176
mar/90	35,00	34,36	3,54	0,189
abr/90	29,92	30,36	3,41	0,030
mai/90	32,94	32,34	3,48	0,111
jun/90	34,85	35,36	3,57	0,226
jul/90	38,05	37,36	3,62	0,296
ago/90	36,53	35,87	3,58	0,244
set/90	30,67	31,12	3,44	0,062
out/90	23,02	22,60	3,12	-0,349
nov/90	19,83	20,12	3,00	-0,499
dez/90	20,48	20,11	3,00	-0,499
jan/91	26,39	25,91	3,25	-0,174
fev/91	23,92	26,00	3,26	-0,169
mar/91	24,73	24,28	3,19	-0,257
abr/91	32,19	32,66	3,49	0,124
mai/91	29,86	29,32	3,38	-0,015
jun/91	28,48	28,90	3,36	-0,034
jul/91	29,41	28,88	3,36	-0,035
ago/91	27,10	26,61	3,28	-0,140
set/91	22,94	23,27	3,15	-0,312
out/91	18,64	18,30	2,91	-0,620
nov/91	20,70	21,00	3,04	-0,444
dez/91	21,90	21,50	3,07	-0,413
jan/92	20,19	19,82	2,99	-0,518
fev/92	29,31	30,76	3,43	0,047
out/95	13,87	13,62	2,61	-1,000
nov/95	15,89	16,12	2,78	-0,783

Mês/Ano	Preço (R\$)	Valor ajustado conforme calendário	Transformação logarítma	Valor normalizado
mai/92	22,61	22,20	3,10	-0,372
jun/92	20,94	21,25	3,06	-0,429
jul/92	24,65	24,20	3,19	-0,261
ago/92	29,56	29,02	3,37	-0,028
set/92	33,57	34,06	3,53	0,178
out/92	33,55	32,94	3,49	0,135
nov/92	29,38	29,81	3,39	0,006
dez/92	29,40	28,87	3,36	-0,035
jan/93	30,48	29,27	3,38	-0,017
fev/93	34,16	37,13	3,61	0,289
mar/93	36,32	35,66	3,57	0,237
abr/93	37,24	37,78	3,63	0,311
mai/93	34,52	33,89	3,52	0,171
jun/93	35,86	36,38	3,59	0,262
jul/93	34,45	33,82	3,52	0,169
ago/93	33,14	32,54	3,48	0,119
set/93	28,04	28,45	3,35	-0,054
out/93	25,28	24,82	3,21	-0,229
nov/93	23,79	24,14	3,18	-0,265
dez/93	24,88	24,43	3,20	-0,249
jan/94	25,61	25,15	3,22	-0,212
fev/94	28,16	30,61	3,42	0,040
mar/94	30,31	29,76	3,39	0,004
abr/94	31,65	32,11	3,47	0,102
mai/94	30,01	29,47	3,38	-0,009
jun/94	30,77	31,22	3,44	0,066
jul/94	35,82	35,17	3,56	0,219
ago/94	34,32	33,70	3,52	0,164
set/94	28,32	28,73	3,36	-0,041
out/94	23,80	23,37	3,15	-0,306
nov/94	25,67	26,04	3,26	-0,167
dez/94	29,76	29,22	3,37	-0,019
jan/95	22,17	21,77	3,08	-0,398
fev/95	21,35	23,21	3,14	-0,315
mar/95	18,78	18,44	2,91	-0,611
abr/95	16,20	16,44	2,80	-0,758
mai/95	14,55	14,29	2,66	-0,939
jun/95	15,98	16,21	2,79	-0,776
jul/95	14,84	14,57	2,68	-0,913
ago/95	16,95	16,64	2,81	-0,742
set/95	16,95	16,64	2,81	-0,742
out/95	14,05	14,25	2,66	-0,941
nov/95	13,87	13,62	2,61	-1,000
dez/95	15,89	16,12	2,78	-0,783
jan/96	17,75	17,43	2,86	-0,683
fev/96	16,22	15,93	2,77	-0,799
mar/96	23,59	23,16	3,14	-0,318

Mês/Ano	Preço (R\$)	Valor ajustado conforme calendário	Transformação logarítma	Valor normalizado
dez/95	17,75	17,43	2,86	-0,683
jan/96	16,22	15,93	2,77	-0,799
abr/96	26,95	27,34	3,31	-0,105
mai/96	27,14	26,65	3,28	-0,138
jun/96	29,89	30,33	3,41	0,028
jul/96	30,20	29,65	3,39	0,000
ago/96	29,97	29,43	3,38	-0,010
set/96	23,23	23,57	3,16	-0,295
out/96	19,68	19,32	2,96	-0,551
nov/96	15,66	15,89	2,77	-0,802
dez/96	19,80	19,44	2,97	-0,543
jan/97	20,94	20,56	3,02	-0,471
fev/97	24,89	27,06	3,30	-0,118
mar/97	28,04	27,53	3,32	-0,096
abr/97	25,22	25,59	3,24	-0,190
mai/97	23,43	23,00	3,14	-0,327
jun/97	25,66	26,03	3,26	-0,168
jul/97	26,14	25,67	3,25	-0,186
ago/97	23,20	22,78	3,13	-0,339
set/97	21,90	22,22	3,10	-0,371
out/97	18,97	18,63	2,92	-0,598
nov/97	14,26	14,47	2,67	-0,922
dez/97	17,10	16,79	2,82	-0,731
jan/98	16,44	16,14	2,78	-0,782
fev/98	18,19	19,77	2,98	-0,521
mar/98	21,98	21,58	3,07	-0,409
abr/98	22,80	23,13	3,14	-0,319
mai/98	25,13	24,67	3,21	-0,237
jun/98	23,44	23,78	3,17	-0,284
jul/98	24,74	24,29	3,19	-0,257
ago/98	21,31	20,92	3,04	-0,448
set/98	19,71	20,00	3,00	-0,506
out/98	20,41	20,04	3,00	-0,504
nov/98	20,39	20,69	3,03	-0,463
dez/98	18,15	17,82	2,88	-0,655
jan/99	16,12	15,83	2,76	-0,807
fev/99	17,08	18,57	2,92	-0,602
mar/99	20,29	19,92	2,99	-0,511
abr/99	21,01	21,32	3,06	-0,424
mai/99	23,30	22,88	3,13	-0,334
jun/99	21,57	21,88	3,09	-0,391
jul/99	22,32	21,91	3,09	-0,389
ago/99	19,41	19,06	2,95	-0,568
set/99	18,60	18,87	2,94	-0,581
out/99	18,42	18,09	2,90	-0,636
nov/99	20,45	20,75	3,03	-0,459
dez/99	20,55	20,18	3,00	-0,495

Tabela B.2 - Conjunto de treinamento da rede

Neurônio													
1	2	3	4	5	6	7	8	9	10	11	12	13	T
0,445	0,593	0,323	0,155	0,146	0,370	0,311	0,497	0,530	0,298	0,513	0,546	0,121	0,593
0,593	0,323	0,155	0,146	0,370	0,311	0,497	0,530	0,298	0,513	0,546	0,121	0,532	0,323
0,323	0,155	0,146	0,370	0,311	0,497	0,530	0,298	0,513	0,546	0,121	0,532	0,748	0,155
0,155	0,146	0,370	0,311	0,497	0,530	0,298	0,513	0,546	0,121	0,532	0,748	0,769	0,146
0,146	0,370	0,311	0,497	0,530	0,298	0,513	0,546	0,121	0,532	0,748	0,769	0,710	0,370
0,370	0,311	0,497	0,530	0,298	0,513	0,546	0,121	0,532	0,748	0,769	0,710	0,750	0,311
0,311	0,497	0,530	0,298	0,513	0,546	0,121	0,532	0,748	0,769	0,710	0,750	0,708	0,497
0,497	0,530	0,298	0,513	0,546	0,121	0,532	0,748	0,769	0,710	0,750	0,708	0,702	0,530
0,530	0,298	0,513	0,546	0,121	0,532	0,748	0,769	0,710	0,750	0,708	0,702	0,728	0,298
0,298	0,513	0,546	0,121	0,532	0,748	0,769	0,710	0,750	0,708	0,702	0,728	0,696	0,513
0,513	0,546	0,121	0,532	0,748	0,769	0,710	0,750	0,708	0,702	0,728	0,696	0,774	0,546
0,546	0,121	0,532	0,748	0,769	0,710	0,750	0,708	0,702	0,728	0,696	0,774	0,685	0,121
0,121	0,532	0,748	0,769	0,710	0,750	0,708	0,702	0,728	0,696	0,774	0,685	0,705	0,532
0,532	0,748	0,769	0,710	0,750	0,708	0,702	0,728	0,696	0,774	0,685	0,705	0,926	0,748
0,748	0,769	0,710	0,750	0,708	0,702	0,728	0,696	0,774	0,685	0,705	0,926	0,699	0,769
0,769	0,710	0,750	0,708	0,702	0,728	0,696	0,774	0,685	0,705	0,926	0,699	0,522	0,710
0,710	0,750	0,708	0,702	0,728	0,696	0,774	0,685	0,705	0,926	0,699	0,522	0,512	0,750
0,750	0,708	0,702	0,728	0,696	0,774	0,685	0,705	0,926	0,699	0,522	0,512	0,372	0,708
0,708	0,702	0,728	0,696	0,774	0,685	0,705	0,926	0,699	0,522	0,512	0,372	-0,385	0,702
0,702	0,728	0,696	0,774	0,685	0,705	0,926	0,699	0,522	0,512	0,372	-0,385	-0,298	0,728
0,728	0,696	0,774	0,685	0,705	0,926	0,699	0,522	0,512	0,372	-0,385	-0,298	-0,166	0,696
0,696	0,774	0,685	0,705	0,926	0,699	0,522	0,512	0,372	-0,385	-0,298	-0,166	0,008	0,774
0,774	0,685	0,705	0,926	0,699	0,522	0,512	0,372	-0,385	-0,298	-0,166	0,008	-0,188	0,685
0,685	0,705	0,926	0,699	0,522	0,512	0,372	-0,385	-0,298	-0,166	0,008	-0,188	-0,082	0,705
0,705	0,926	0,699	0,522	0,512	0,372	-0,385	-0,298	-0,166	0,008	-0,188	-0,082	-0,240	0,926
0,926	0,699	0,522	0,512	0,372	-0,385	-0,298	-0,166	0,008	-0,188	-0,082	-0,240	0,224	0,699
0,699	0,522	0,512	0,372	-0,385	-0,298	-0,166	0,008	-0,188	-0,082	-0,240	0,224	0,379	0,522
0,522	0,512	0,372	-0,385	-0,298	-0,166	0,008	-0,188	-0,082	-0,240	0,224	0,379	0,499	0,512
0,512	0,372	-0,385	-0,298	-0,166	0,008	-0,188	-0,082	-0,240	0,224	0,379	0,499	0,335	0,372
0,372	-0,385	-0,298	-0,166	0,008	-0,188	-0,082	-0,240	0,224	0,379	0,499	0,335	0,105	-0,385
-0,385	-0,298	-0,166	0,008	-0,188	-0,082	-0,240	0,224	0,379	0,499	0,335	0,105	0,250	-0,298
-0,298	-0,166	0,008	-0,188	-0,082	-0,240	0,224	0,379	0,499	0,335	0,105	0,250	0,462	-0,166
-0,166	0,008	-0,188	-0,082	-0,240	0,224	0,379	0,499	0,335	0,105	0,250	0,462	-0,027	0,008
0,008	-0,188	-0,082	-0,240	0,224	0,379	0,499	0,335	0,105	0,250	0,462	-0,027	0,108	-0,188
-0,188	-0,082	-0,240	0,224	0,379	0,499	0,335	0,105	0,250	0,462	-0,027	0,108	0,263	-0,082
-0,082	-0,240	0,224	0,379	0,499	0,335	0,105	0,250	0,462	-0,027	0,108	0,263	0,226	-0,240
-0,240	0,224	0,379	0,499	0,335	0,105	0,250	0,462	-0,027	0,108	0,263	0,226	0,190	0,224
0,224	0,379	0,499	0,335	0,105	0,250	0,462	-0,027	0,108	0,263	0,226	0,190	0,348	0,379
0,379	0,499	0,335	0,105	0,250	0,462	-0,027	0,108	0,263	0,226	0,190	0,348	0,304	0,499
0,499	0,335	0,105	0,250	0,462	-0,027	0,108	0,263	0,226	0,190	0,348	0,304	0,611	0,335
0,335	0,105	0,250	0,462	-0,027	0,108	0,263	0,226	0,190	0,348	0,304	0,611	0,776	0,105
0,105	0,250	0,462	-0,027	0,108	0,263	0,226	0,190	0,348	0,304	0,611	0,776	1,000	0,250
0,250	0,462	-0,027	0,108	0,263	0,226	0,190	0,348	0,304	0,611	0,776	1,000	0,449	0,462
0,462	-0,027	0,108	0,263	0,226	0,190	0,348	0,304	0,611	0,776	1,000	0,449	0,591	-0,027
-0,027	0,108	0,263	0,226	0,190	0,348	0,304	0,611	0,776	1,000	0,449	0,591	-0,188	0,108
0,108	0,263	0,226	0,190	0,348	0,304	0,611	0,776	1,000	0,449	0,591	-0,188	-0,543	0,263
0,263	0,226	0,190	0,348	0,304	0,611	0,776	1,000	0,449	0,591	-0,188	-0,543	-0,453	0,226

